

Banking Competition, Market Structure and Financial Stability of the Gulf Cooperation Council Region

Sara Alfaihani

The thesis is submitted in partial fulfilment of the requirements for
the award of the degree of

DOCTOR OF PHILOSOPHY

of the University of Portsmouth



Portsmouth Business School

Portsmouth, United Kingdom

March 2020

Abstract

The first introduction of market competition concept was by Adam Smith in the 18th century. This inspired scholars for two main views of competition: static and dynamic. The development of several theories of competition has resulted in numerous methodologies to measure banking competition that can be classified into two main categories: structural and non-structural. Moreover, the

history of competition law began in the USA, UK and the EU who have succeeded in building solid antitrust regimes. On the other hand, the young economies of the Gulf region have enacted competition law very recently and lacks clear merger guidelines especially for the banking sector. A comprehensive and detailed review of competition theories, policies, and measures applied by both academics and antitrust authorities worldwide show that there is a disagreement in the banking literature regarding which competition measures best reflect the degree of competition given the strengths and drawbacks of each measure. Also, banking regulations in the Gulf Cooperation Council (GCC) countries lacks clear merger guidelines and ignores measures of competition as one of the criteria. This thesis proceeds with three main empirical analyses.

The first empirical chapter explores the market structure and the size-structure relationship in the GCC banking industry over the period 2000-2017. The concentration measures indicate that the GCC banking market are operating under somewhat concentrated conditions except for UAE and Saudi Arabia. Also, there exist a lower bound to concentration and the market remains concentrated regardless of market size. This implies that the banking industry of the GCC is an endogenous sunk cost industry, whereby the limiting levels of concentration are positive and bounded well above zero. Moreover, the limiting levels are increasing over time, so largest banks become more dominant, and are different between individual countries due to market size and set-up costs. Finally, the values of the limiting and actual levels of concentration are astonishingly close, which indicates that the banking markets of the GCC countries are operating under a long-run equilibrium.

The second empirical chapter investigates the competition in the GCC banking sector by applying the modified Lerner Index that takes the scale of operation

ii

into account. We find evidence that up to 86% of banks are optimally scaled. About 10% of banks are larger than the maximum productive scale size, and the market power of nearly all of them is larger than that suggested by the conventional Lerner Index. Half of the banks that operate with economies of scale actually have larger market power than the conventional Lerner Index predicts. These results may be manifestations of the future, where some banks that have large market power consolidate and grow even further.

The third empirical chapter examines the relationship between banking competition and financial stability in the GCC banking sector over the period 2000–2017. The paper proceeds with three main analyses. First, we examine the scale economies of GCC banks in a statistical sense and find that around 85% of bank-years are operating under economies of scale whereas the remaining 15% are at the optimal scale. Second, we measure banks' market power using the Lerner Index and a scale-corrected Lerner Index. Both versions of the Lerner Index indicate that banks under economies of scale enjoy higher market power than banks with the optimal scale. Finally, we examine the competition–stability relationship between GCC banks. Overall, higher market power leads to greater financial stability, in line with the competition–fragility view. This full-sample relationship holds across all GCC countries except for Qatar, where the cross country analysis shows that higher competition contributes positively to stability.

Keywords: GCC; Competition; Retail Banks; Scale Economies; Concentration; Competition Law; Competition Theory; Bank Competition; Financial Stability

JEL Classification: D4; G2; C5; L1

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word Count: 71,850

iv

Contents

List of Figures..... xi **List of**

Tables xiii

Abbreviationsxv

Acknowledgments	xvi
Chapter 1. Introduction.....	1
1. Philosophy of Science	2
Motivations of Thesis	4
Research Questions.....	5
4. Aims and Objectives	6
5. Contributions.....	7
5.1. First Empirical Chapter: Chapter 4.....	8
5.2. Second Empirical Chapter: Chapter 5	9
5.3. Third Empirical Chapter: Chapter 6	9
6. Linkage among the Empirical Chapters.....	10
Structure of Thesis	11
Chapter 2. Measuring Banking Competition: A Critical Review of Theories and Methodologies	14
1. Introduction.....	14
Review Method.....	20
3. Competition in Economic Theory.....	22
3.1. Static Theories of Competition.....	23
3.2. Dynamic Theories of Competition	31
v	
4. Structural Approaches (EIO Theory)	34
4.1. Non-formal Approaches	34
4.2. Formal Approaches.....	37
5. Non-structural Approaches (NEIO Theory)	42
5.1. First Generation of Non-structural Approaches.....	44

5.1.1 Lerner Index (Lerner, 1934).....	44
5.1.2 Conjectural Variation Model (Iwata, 1974).....	47
5.1.3 Panzar-Rosse H-statistic (Rosse & Panzar, 1977).....	49
5.1.4 Bresnahan-Lau (1982) Mark-up Model.....	53
5.2. Second Generation of Non-structural Approaches.....	57
5.2.1 Persistence of Profits Model (Mueller, 1977, 1986)	57
5.2.2 Hall-Roeger Model (Hall, 1988; Roeger, 1995).....	59
5.2.3 Boone Indicator (Boone, 2008).....	61
5.2.4 Competition Efficiency Frontier (Bolt & Humphrey, 2010).....	64
5.2.5 Distance to the Lower Bound to Concentration Approach (Tabacco, 2013).....	67
6. Conclusion	69
6.1. Conclusion	69
6.2. Future Trends.....	71

Chapter 3. Competition Policy and its Role in the Banking Sector: An International Comparison73

1. Introduction.....	73
Review Method.....	77
Antitrust Policy of the USA.....	79
Competition Policy of the UK.....	87
Competition Policy of the EU	91
6. Competition Policy of the GCC	96
6.1. Saudi Arabia.....	97
6.2. Qatar	100
6.3. Kuwait.....	102
6.4. United Arab Emirates.....	103
6.5. Oman	106
6.6. Bahrain	107

7. Competition Policy of the Rest of the World.....	110	7.1.
Canada.....	111	
7.2.		Australia
.....	112	7.3.
Russia	112	
7.4.		
India.....	113	
7.5.		
Pakistan.....	113	
7.6.		Japan
.....	114	7.7.
China	114	
7.8.		Latin
America.....	115	
8. Competition Policy of the World Bank	117	9.
Evidence of Effective Competition Law in the Banking Sector	119	
10. Conclusion.....	121	10.1.
Conclusion.....	121	
10.2.		Future
Trends.....	123	
 Chapter 4. A Stochastic Frontier Approach to Modelling Market		
Structure in the Banking Industry: An Application to the Gulf		
Cooperation Council (GCC)	124	
1. Introduction.....	124	2.
Literature Review	128	3. Data
.....	132	
4. Methodology	135	4.1.
Market Structure	135	
4.2. The Lower Bound to Concentration		
.....	136	
5. Empirical Findings.....	138	5.1.
Market Structure	138	

5.2.	Size–structure
Relationship.....	142 5.2.1.
Baseline Model.....	142
5.2.2. Extension 1:	Cross-time
heterogeneity.....	148 5.2.3. Extension 2: Cross-
country heterogeneity	152 5.2.4. Extension 3:
Time and country heterogeneity.....	156
6. Discussion	161 7.
Conclusion	163
 Chapter 5. Banking Competition of the GCC: The Tale of Scale	
Economies	165
1. Introduction.....	165 2. The
Banking Sector in the GCC.....	169 3. Literature
Review	172
4. Data and Methodology	177 4.1.
Data.....	177
4.2. Econometric	Methodology
.....	183
5. Empirical Results and Discussion	185 5.1.
Economies of Scale.....	185
5.2. Competition	
.....	191 5.3.
Discussion	193
6. Conclusion	195
 Chapter 6. Scale, Competition, and Financial Stability in GCC	
Banking.....	197
1. Introduction.....	197
2. Data and Sample Construction.....	201 2.1.
Data Sources.....	201
2.2. Sample	

Construction.....	202
3. Scale Economies, Banking Competition, and Financial Stability.....	203
viii	
3.1. Scale Economies.....	204
3.1.1. Method	204
3.1.2. Sample Characteristics	207
3.1.3. Results.....	209
3.2. Competition	209
3.2.1. Lerner Index	210
3.2.2. Scale-Corrected Lerner Index	210
Results.....	211
3.3. Financial Stability and Competition.....	214
Empirical Strategy.....	214
Sample Characteristics	218
Results.....	220
Extension: Cross-Country Heterogeneity.....	228
Robustness Check.....	231
4. Competition-Stability Nexus in the GCC Countries	235
Conclusion	237
7. Conclusion	239
1. First Empirical Chapter:.....	239
A stochastic frontier approach to modelling market structure in the banking industry: An application to the Gulf Cooperation Council (GCC).....	
1.1. Findings.....	239

1.2. Policy Implications	240
2. Second Empirical Chapter:	240
Banking competition of the GCC: The tale of scale economies	240
Findings.....	240
2.2. Policy Implications	241
3. Third Empirical Chapter:	242
Scale, competition, and financial stability in GCC banking	242
ix	
3.1. Findings.....	242
3.2. Policy Implications	242
4. Synthesis.....	243
Limitations.....	243
Direction for Future Studies.....	244
References.....	246
Appendices.....	269

x
List of Figures

Figure 2.1. Methodologies of measuring banking competition	
19 Figure 2.2. The number of published papers on competition in banking (1980-2017).....	
20 Figure 2.3. Extraction and interpretation of the PR H-statistic.	
50 Figure 3.1. First competition law enactment.....	74
Figure 3.2. The influence of USA & EU competition laws over countries around the world	75
Figure 3.3. First competition law in the rest of the World.....	111
Figure 4.1. Time series evolution of GCC banking markets.	127
Figure 4.2. Time series evolution of concentration in the GCC banking markets.....	
141 Figure 4.3. Concentration and market size in the GCC banking markets. . 145	
Figure 4.4. Lower bound to concentration of the baseline model.....	146
Figure 4.5. Lower bound to concentration of Extension 1.	150
Figure 4.6. The limiting level of concentration across time.	151
Figure 4.7. The limiting level of concentration across countries.	154
Figure 4.8. Lower bound to concentration of Extension 2.	155
Figure 4.9. The limiting level of concentration across time and countries.. 159	
Figure 4.10. Lower bound to concentration of Extension 3.	160
Figure 5.1. Time series evolution of GDP 2000-2018	167
Figure 5.2. Size of GCC national retail banking markets 2000–2017	171
Figure 6.1. Time series evolution of bank market power for the GCC countries' banking sector over the period 2000–2017.....	213

Figure 6.2. Time series evolution of bank market power, financial stability,
and profit volatility.

221 **Figure 6.3.** The degree of the competition–stability relationship across
GCC countries.

..... 230

List of Tables

Table 1.1.	Structure of thesis	12
Table 2.1.	Distribution of articles by knowledge area and journal.....	21
Table 3.1.	Distribution of articles by knowledge area and journal.....	78
Table 4.1	Summary statistics.....	134
Table 4.2	Banking market structure and size of the GCC.....	140
Table 4.3	Baseline model estimation results.	147
Table 4.4	Extension 1 estimation results.	149
Table 4.5	Extension 2 estimation results.	153
Table 4.6	Extension 3 estimation results.....	158
Table 5.1.	Market structure indices of GCC banking sectors in 2017	172
Table 5.2.	Number of bank-year observations per country	179
Table 5.3.	Data description.....	180
Table 5.4.	Sample summary statistics.....	182
Table 5.5.	Empirical estimation of translog cost function.....	188
Table 5.6.	Empirical estimation of Lerner Index.....	190
Table 5.7.	Empirical estimation results.....	190
Table 6.1	Summary statistics.....	208
Table 6.2	Bank-year distribution of scale economies.	209
Table 6.3	The Lerner Index and other characteristics of competition in the GCC countries.	213
Table 6.4	Empirical estimation of the Lerner indices.....	214
Table 6.5	Summary statistics of the competition–stability model.	219
Table 6.6	Competition-stability relationship: Full sample regression.....	226
Table 6.7	Competition-stability relationship.....	229

Table 6.8 Competition–stability relationship with the scale-corrected Lerner Index as a measure of competition	
233	

xiv
Abbreviations

GCC – Gulf Cooperation Council
UAE – United Arab Emirates
USA – United States of America
UK – United Kingdom
EU – European Union
EIO – Empirical Industrial Organisation NEIO
– New Empirical Industrial Organisation MP –
Market Power
SCP – Structure- Conduct-Performance
RMP – Relative-Market-Power
ES – Efficient Structure
ESX – X-efficiency Efficient Structure Hypothesis
ESS – Scale-efficiency Efficient Structure
Hypothesis DOJ – Department of Justice
SF – Stochastic Frontier
 CR_k – K-bank Concentration Ratio
HHI – Hefindahl-Hirschman Index
M&A – Mergers and Acquisitions
R&D – Research and Development

Acknowledgments

Undertaking this PhD has been a genuinely extraordinary encounter for me, and it would not have been conceivable to manage without the help and direction that I got from numerous individuals.

Firstly, sincere gratitude and thankfulness goes to my supervisors *Prof. Shabbar Jaffry* and *Dr Oleg Badunenko* for the continuous support of my PhD study and related research, for their patience, inspiration, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisors and mentors for my PhD study.

I gratefully acknowledge my employer, *University of Bahrain* for the funding received towards my PhD. This PhD study would not have been possible without their generous sponsorship.

My deep gratefulness goes to His Excellency *Shaikh Fawaz bin Mohammed Al Khalifa*, the Ambassador to United Kingdom for the continuous support and patronage to all Bahraini students in the U.K. I also greatly appreciate the support and assistance received from the Cultural Counsellor, *Mr Manea Al Manea* at the Bahraini Cultural Attaché Office.

Many thanks to *Prof. Laura Spierdijk* for the guidance and valuable remarks towards my PhD study.

A huge thank you to my dearest friends back home, *Aysha Fakhroo, Amina Al Mannai, Bushra and Aysha Aljassmi* for the unending support they have

provided even though from a distance. A special thank you to the kindest friends from Saudi Arabia *Al-Wuhaib* family for the support and wonderful time we have had in Portsmouth. I would also like to say thank you to the incredible friends from Greece, *Marianna Daskalapolou and Panagiotis Tzouvanas* for the support and exceptional memories we created.

A special thanks goes to *Dr Fraser Cumming* at Southampton General Hospital for the tremendous care, support and spirit lifting during my time in England. I cannot speak highly enough of the appreciation I give towards the high-quality health services and ethics in this hospital.

Last but not the least, I would also like to say a heartfelt thank you to my parents and to my siblings, *Abdulwahab, Fahad, Mona and Noora* for supporting me spiritually throughout writing this thesis and encouraging me to follow my dreams.

Chapter 1. Introduction

The world economy has been facing persistently low growth, which has brought more universal attention to competition as a substantial driver of performance and superb innovation in a national economy. Hence, a driver of economic growth and consumer welfare. Competition in the marketplace will foster economic growth and prosperity by promoting the productivity of firms and industries, enhancing the competitiveness of domestic firms internationally (which will lead to higher export levels), and ensuring that lower-income consumers benefit from vibrant economic activity (Goodwin & Martinez Licetti, 2016).

The banking sector is a major area of interest within the field of competition, because of the crucial role that the financial intermediation services play in the economy. The context of the recent financial crisis has also ignited this interest (e.g. Efthyvoulou & Yildirim, 2014). Of particular concern is the macroeconomic outcomes of banking competition (e.g. Cetorelli, 2004; Coccoresse, 2004). Competition within banking is a fundamental factor of promoting three main aspects of this sector: efficiency, access to finance, and systemic financial stability. The relationship between banking competition and these three aspects has been explored in several studies. While the literature yields mixed results, most research suggests that banking competition enriches the banks' efficiency and improves access to credit without necessarily undermining the stability of the financial system (Boyd & De Nicoló, 2005; World Bank, 2012; Beck et al., 2013; Schaeck & Cihák, 2014; Kick & Prieto, 2015).

The Gulf Cooperation Council (GCC) region offers an interesting case of analysis in this field, since all the GCC governments are pursuing economic diversification policies to reduce heavy reliance on oil as a primary source of income. Henceforth the Gulf region and GCC will be used interchangeably. The banking sector, which dominates the financial sector, is seen as a crucial sector of the region and has been a key driver for the GCC's economies and an important factor in achieving these plans (Abuzayed et al., 2018). Hence, it is crucial to assess the degree of

associated with the market structure. However, to date, there are only a limited number of studies conducted for the Gulf region (e.g. Al-Muharrami et al., 2006).

Moreover, competition policy generally comprises a set of policies and provisions that maintain fair competition in the marketplace and ensure that there are no restrictions to competition that will reduce economic welfare. The chief objective of competition policy is thus to promote economic growth and social prosperity by protecting competition (Goodwin & Martinez Licetti, 2016). Competition policy and antitrust laws and regulations have a long history that can be traced as far back as the 19th century. Yet again, the GCC countries have not paid attention to competition law and policy until very recently. This attention is mainly driven by the economic visions of these countries of achieving a diversified economy and reducing the heavy reliance on oil revenues.

To date, empirical analysis to test the degree of competitiveness in the GCC banking sector are exceptionally limited and are applying traditional methods of measuring competition that are criticised for many shortcomings. Besides, the existing literature in that field for the GCC countries is relatively not recent and did not take into account the recent global financial crisis period. This might have had a significant influence on the market structure of these six countries. Analysis of the relationship between the competition and financial stability in the Gulf region is also relatively sparse.

1. Philosophy of Science

This section presents the research approach philosophy of the thesis prior to further enlightenments about the market structure, competition and financial stability in the banking industry.

A well-known quote of the Greek philosopher Heraclitus (535–475 B.C) regarding the everlasting changes in the universe “there is nothing permanent except change” (Wheelwright, 1974). In the banking sector context, perpetual changes such as global crises and expansion of market size influence the decision

questions. Is there a relationship between the market structure and market size? Does competitive behaviour diverge among the same sector in different geographical locations? What is the role of scale economies? What are the effects of these changes on the stability of the financial sector? Amongst others, how these changes are reflected on the oil based GCC economies? In this thesis, these philosophical considerations have been investigated through empirical examinations.

The research philosophy of this thesis is conducted according to the epistemology approach with positivism doctrine. The epistemology approach underlies the entire research design and governs the selected theoretical doctrine (i.e. positivism or interpretivism). It is primarily about “how we know what we know” (Crotty, 1998, p. 8). This approach is mainly based on assessing the economic theories through the collection of real-world data. The positivism doctrine highlights that the researchers should take an objective position and should not be driven by their own preferences. One way to satisfy this assumption is by primarily considering and testing the existing theoretical framework and then generating new principles. In other words, the positivism doctrine relates more to testing existing theories rather than developing new ones (Crotty, 1998). In this framework, the empirical chapters of this thesis are constructed by, first deriving the research hypotheses, then collecting the data and finally using the appropriate methodology.

A quantitative methodology of econometrics with secondary data from credible sources is adopted in this thesis. The empirical methods have been selected in adherence with previous relevant literature and have been checked for robustness. Subsequently, the selected variables exhibit both the theoretical predictions as well as suggestions from previous studies. The chosen methodology of econometrics is in line with the positivism doctrine. Under this method, inferences are produced using more a deductive approach (theory driven) and less an inductive approach (data driven). The inductive approach has been only utilised to link the findings of the thesis with previous empirical

2. Motivations of Thesis

The GCC economies have experienced a remarkable and continuous growth over the last two decades. According to the World Bank GDP ranking of 2018, the GCC economies are ranked jointly as the top 12th economy worldwide. The young yet vibrant economies of the GCC, have benefited from the massive oil revenues which have contributed highly in boosting the economy of the region. As a result, the GCC economy has started to compete globally and has already overtaken some developed countries such as Switzerland.

During the last three decades, the GCC banking markets went through significant and promising reforms. Primarily, these reforms introduced the implementation of financial liberalisation policies and financial restructuring policies. More recently, the central banks in the GCC countries have introduced corporate governance-related standards and risk management rules in accordance with Basel III. The sole objective is to enhance the sector's competitiveness (Abuzayed et al., 2018). Banks do and will be playing a prominent role in GCC countries, so special attention needs to be paid to the market structure, competitiveness' and stability of the banking industry in the Gulf region. Even though there is well established literature of competition and stability, there is still a large unexplored part especially for the GCC countries. Therefore, this thesis aims to fill these crucial gaps in the literature.

This research makes a number of important methodological and empirical contributions to the literature. *First*, the traditional inverse relationship between market structure and market size has been criticised for only being valid in certain industries. Sutton (1991) argues that endogenous fixed costs play a dominant role in determining the equilibrium structure and in such case the industry remains concentrated regardless of the size. Hence, the inverse size structure relationship breaks down. Surprisingly, the literature does not provide

much analysis in this field. For these reasons, Chapter 4 of this thesis extensively investigates the size-structure relationship in the GCC banking industry.

Second, there is a significant debate in the academic literature concerning the reliability of the different measures of competition. While the Lerner Index has come to represent the standard measure of market power (Blair & Sokol, 2014), it fails to recognise the actual market power in the presence of scale economies (Spierdijka & Zaourasa, 2018). In this context, we apply both the traditional and the modified Lerner Index, that takes the scale of operations into account in order to explore the difference of competitiveness between banks under economies of scale and diseconomies of scale (Chapter 5).

Third, the literature provides no consensus about the relationship between bank competition and financial stability. The competition–fragility view, on the one hand, suggests that higher competition leads to higher fragility (Keeley, 1990). On the other hand, the competition–stability view argues that more competition results in greater stability (Boyd & De Nicoló, 2005). In Chapter 6, therefore, we investigate the impact of the bank competition on the bank stability of the Gulf region, as well as across individual GCC countries.

Finally, the GCC authorities until very recently, have not paid attention to competition law and policy. The findings of this thesis can serve as solid guidelines for policymakers. It provides a comprehensive review of competition measures applied by authorities worldwide. Also, we draw attention to neglected issues in banking, that can have significant effects on the financial stability and consumer welfare.

3. Research Questions

In order to remedy the gap in the literature, this thesis presents three empirical examinations. The research questions that this thesis attempts to posit are the following:

- i. Is there a difference between measures of competition applied by academics and by competition authorities?
- ii. Why studying market structure is essential for analysing competitive conditions?
- iii. What role scale economies play in the analysis of competition?
- iv. Does

banking competition enhances financial stability in the Gulf region? v. What policy implications and competition legislations improvements are needed in the Gulf region?

4. Aims and Objectives

Having discussed about the motivations of the thesis, we now turn to list the aims and objectives. This thesis presents two comprehensive reviews of literature chapters and three empirical examinations and, hence, the thesis has six aims and several objectives.

The *first* aim is to provide a comprehensive and detailed review of competition theories and measures, with a focus on the banking sector (Chapter 2). The *second* aim is to review the history of competition policy and competition measures applied by authorities worldwide (Chapter 3). Therefore, this study critically reviews both the academic and the practical literature concerning competition in the banking sector and evaluates the most significant features within the field. The *third* aim of the thesis is to investigate the size-structure relationship in the GCC banking industry. The *fourth* aim of the study is to examine the competitive conditions of GCC banking sector using most recent methodologies. The *fifth* aim is to identify the key influences of banking competition on financial stability in the GCC. The *final* aim is to interpret the results and propose recommendations and policy implications to the regulatory authorities and participants in the banking sector of the Gulf region. It is hoped, this will be the first compressive study of competition in the banking sector of the GCC region.

The objective of the first empirical examination (Chapter 4) is to investigate the current market structure and the size-structure relationship for the banking industry of the GCC during the period 2000-2017. In Particular, we measure the current concentration levels in the GCC banking sectors, and investigate whether

larger banking markets exhibit unconcentrated conditions compared to their smaller counterparts. The existing literature has not paid much attention to this field of research. We address this issue and investigate whether the inverse size structure relationship applies to the banking industry of the Gulf

region.

The objective of Chapter 5 is to undertake research to examine the degree of competition in the banking industry of GCC countries, analysing 79 banks for the period 2000-2017. We utilise the Lerner Index and a scale corrected Lerner Index, in order to have precise estimates of banking competition. By considering the scale economies, we provide estimates of competition for banks in the different stages of the production cycle.

Finally, yet just as important, the third empirical chapter (Chapter 6) intends to examine if banking competition is beneficial to the stability of the financial system. Given that the competition can either enhance or undermine stability, we investigate whether competition in the banking system of the GCC leads to a higher stability or a higher fragility.

5. Contributions

This research makes valuable contributions. The main objective of this study is to provide the participant in the banking sector, managers, researchers and policymakers with an overview of the existing literature regarding the competition in banking. In addition, the competition research for the GCC countries is rather limited, the majority of the studies have been conducted for developed economies. Hence, we fill this important yet unexplored part of the literature with new empirical evidence. Even though the thesis has a few conceptual and theoretical contributions, the main contributions are empirically related. In particular, we provide empirical evidence regarding market structure, competition and financial stability in an area which is drawing attention to the banking industry, the GCC region.

For all the empirical chapters, this thesis makes an original contribution to the GCC banking sector through the data preparation, methodological and empirical

contributions, as these represent the basis for accurate empirical results. In this context, tremendous effort has been made to compile a clean and cross-validated dataset of the GCC banks for the last 18 years. The dataset includes *all* national

retail banks operating in the GCC. The sample includes longitudinal data for the period of 2000–2017. The earlier years were excluded due to the unavailability of data and major changes in the accounting standards. This comprehensive dataset comes from four different sources: Bloomberg; annual reports of banks; S&P Global Market Intelligence; and BankScope databases. This dataset has undergone several processes to assure reliability and consistency over the entire sample period. This dataset is used for the three empirical examinations and represent a major contribution to the empirical analysis.

The empirical work presented here, provides the most comprehensive study on the topic of banking competition and can serve as a reference point not just for the GCC countries, but for all various banking sectors in the world. Moreover, the empirical and methodological contributions postulate a clear and well established guidance towards research in the competition literature for other industries as well. Hence, this thesis contributes to the extensive literature of competition both empirically and methodologically and is not limited to a specific region. The specific contributions of each empirical chapter are presented below.

5.1. First Empirical Chapter: Chapter 4

In the first empirical chapter, the study goes beyond the traditional literature and allows the market structure to vary over time and across countries. To the best of our knowledge, this is the first study to test Sutton (1991) theory in the GCC banking industry.

In addition, this is the first attempt to evaluate the market structure-size relationship with a time-varying stochastic frontier approach. Particularly, the Weibull distribution using maximum likelihood and minimum distance estimators are employed (Dick, 2007; Ellickson, 2007). The literature, moreover, utilises cross-sectional data and estimates the equilibrium market structure at

one point of time. In this chapter, we utilise panel data and extend the original model to capture any heterogeneity over time and across individual countries.

5.2. Second Empirical Chapter: Chapter 5

The second empirical chapter contributes to the growing area of banking competition in four ways. *First*, this research examines the degree of banking competition in the six countries of the Gulf region, who pursue major diversification policies, and pay special attention to the banking sector as a key driver for such policies. It is therefore of great importance, to assess the competitiveness of the banking market in this region, about which the existing literature does not provide much evidence. *Second*, we apply the modified Lerner Index, which provides a more comprehensive picture of competition and explores the competitive differences between banks operating under economies of scale and diseconomies of scale. *Third*, a tremendous effort has been made to compile a clean and cross-validated dataset of GCC banks for the last 18 years. The dataset includes *all* national retail banks operating in the GCC. This dataset has gone through several processes to assure reliability and consistency over the entire sample period. *Finally*, this study offers empirical evidence that can serve as a solid base for policy implications.

5.3. Third Empirical Chapter: Chapter 6

This third empirical chapter contributes to the rich literature on the competition–stability relationship in several ways. *First*, we examine the impact of competition on the stability of the GCC banking sector. The GCC governments have been following an enormous economic diversification policy to reduce their heavy reliance on oil as their primary source of income. The justifiable concern about economic development is therefore the competition and market power that are prevailing in the banking sector and their impact on financial stability. *Second*, this study examines the average competition–stability relationship in the region as well as across countries. It is the first study – to the best of our knowledge – to explore the heterogeneity of this relationship across the GCC countries and hence contributes directly to the ongoing debate on regulatory

reforms. *Third*, this study also tests for the impact of competition on the driving forces of financial stability represented by banks' profits, income volatility, and capital levels. These in turn present the main driver of bank stability in the Gulf region. *Fourth*, we consider the impact of the presence of economies of scale on

both the competition and the financial stability level. This phenomenon has largely been ignored in the competition–stability literature. *Moreover*, we estimate the cost function using system equations and test the scale economies in a statistical sense. *Furthermore*, this study is based on a comprehensive and well-scrutinised dataset of GCC banks that includes all national retail banks in the region. *Lastly*, the findings can have significant implications for policymakers and regulators in the Gulf region and worldwide.

6. Linkage among the Empirical Chapters

Banks in the GCC play a central role in the national diversification plans as the main finance providers of a broad range of sought-after investments in different sectors of the economy. As can be seen from Chapters 2 and 3, however, empirical investigations of the banking markets of this region are relatively sparse. This is the first study in the PhD research domain, – to the best of our knowledge that explores the market structure, competitiveness, and financial stability in the GCC banking systems.

In particular, the thesis brings together three major issues. The first issue, which is presented in the first empirical chapter (Chapter 4), is how concentrated are the banking markets in the GCC, and does concentration levels vary radically with market size. This chapter ascertains that the GCC banking markets in general have been operating under concentrated conditions. Although the level of concentration varies with the size of markets, the markets will always exhibit some concentration and will never reach unconcentrated conditions. This is due to the nature of the banking industry, whereby quality improvements rely heavily on endogenous fixed costs and thereby raise entry barriers. Such industries can be termed as natural oligopoly industries. This provide an interesting conclusion that concentration is not always harmful for consumer

welfare. In this case, for instance, concentration leads to greater a quality of services provided to consumers.

Having examined the market structure which the GCC banking markets are operating under, Chapter 5 provides estimates of the degree of competition.

Overall, the levels of competition lie somewhat in the middle between the two market extremes, monopoly and perfect competition. It can be said the banking markets of the GCC have an oligopolistic nature, which coincides with the findings of the first empirical chapter. Moreover, the ranking of countries in terms of concentration and competition levels varied for some countries. This emphasises the criticism of other scholars, who argue that concentration measures are not always a good proxy for competition (see e.g. Claessens and Laeven 2004; Carbo-Valverde, Rodríguez-Fernández, and Udell 2009). In this chapter, we also test for the scale of operation and find that most banks in the GCC are operating under the optimal return to scale.

Given the concentration and competition levels as well as the scale economies, we utilise all these aspects and test their relationship with the stability of the financial system in the last empirical chapter. While the literature yields inconclusive results, we find evidence that the banking competition in the Gulf region leads to a more fragile banking system. This in turn confirms the previous findings that perfect competition in GCC banking might not be the optimal case for consumer welfare.

7. Structure of Thesis

The remainder of the thesis is structured as shown in Table 1.1. Chapters 2 and 3 provide a comprehensive review of literature. Chapter 2 of the banking competition review looks at the competition theories, methodologies and their applications in the banking sector. Chapter 3 of the competition policy review summarises the competition policy, laws and measures applied by competition and antitrust authorities worldwide. Chapters 4, 5, and 6 are the three empirical investigations adhering to standard empirical papers. Chapter 4 examines the market structure and the size-structure relationship of the GCC banking industry,

whereas Chapter 5 measures the degree of competition. Chapter 6 investigates the competition-financial stability relationship in the GCC banking sectors. The final part of the thesis, Chapter 7, concludes with overall findings, policy implications of results, and limitations of the study.

Table 1.1. Structure of thesis

<i>Chapters</i>	<i>Title of Chapters</i>	<i>Contents of Chapters</i>
Chapter 1	Introduction	Background of the topic, motivations, aims and contributions
Chapter 2	Competition policy and its role in the banking sector: An international comparison	applications in the banking sector. Review of competition policy, laws and measures applied by competition authorities worldwide.
Chapter 3	A stochastic frontier approach to modelling market structure in the banking industry: An application to the Gulf Cooperation Council (GCC)	Estimates of market structure and size-structure relationship.
Chapter 4	Measuring banking competition: A critical review of theories and methodologies	stability in GCC banking Measurement of bank competition using modified measure of competition. Investigation of competition stability relationship in the GCC banking markets.
Chapter 5	Banking competition of the GCC: The tale of scale economies	
Chapter 6	Scale, competition, and financial	
Chapter 7	Conclusion	Main findings, policy implications, limitations, and future directions.

12

13

Chapter 2. Measuring Banking Competition: A Critical Review of Theories and Methodologies

1. Introduction

In recent times, the world economy has been facing persistently low growth, which has brought more universal attention to competition as a substantial driver of performance and superb innovation in a national economy. Hence, a driver of economic growth and consumer welfare. Competition in the marketplace will foster economic growth and prosperity by promoting the productivity of firms and industries, enhancing the competitiveness of domestic firms internationally (which will lead to higher export levels), and ensuring that lower-income consumers benefit from vibrant economic activity (Goodwin & Martinez Licetti, 2016).

This recognition of competition has particular significance in the developing economies, where low growth has been threatening development and shared prosperity. For instance, World Bank Group publications show that a substantial increase of competition in Tunisia is estimated to have increased labour productivity growth by 5% in one year. Besides, it is thought that refining the competition laws of sectors that restrict competition in Brazil, Kenya, Peru and South Africa could add 0.5% to their economic growth. It has also been found that intensifying competition in maize and sugar markets could decrease poverty by 2% in Kenya (Goodwin & Martinez Licetti, 2016).

Competition within banking, moreover, is a fundamental property of promoting three main aspects of this sector: efficiency, access to finance, and systemic financial stability. The relationship between banking competition and these three aspects has been explored in several studies, and while the literature yields mixed results, most research suggests that banking competition enriches banks' efficiency and improves access to credit without necessarily undermining the stability of the financial system. The context of the recent financial crisis has also

14

ignited the interest of academics and policy makers in measuring banking competition and reshaping competition policies (Boyd & De Nicoló, 2005; World Bank, 2012; Beck et al., 2013; Schaeck & Cihák, 2014; Kick & Prieto, 2015).

The concept of market competition was first introduced by Adam Smith in his classic work *The Wealth of Nations*. Smith (1776) used the term competition to represent the doctrine of free trade and *laissez faire*. According to the Smithian insight, free competition and trade without government intervention are vital to long-run equilibrium in any market. Smith viewed monopoly as the opposite of free competition and used it to denote any barrier to free trade. Inspired by Smith, subsequent competition theories have been developed and contribute to two major views of competition.

On the one hand, standard theory refers to competition as a tranquil or static state of equilibrium in which rivals in the marketplace compete until equilibrium is reached. Once such equilibrium holds in the marketplace, there will be no further interaction between rivals and the market will come to rest. However,

powerful critics shed new lights on viewing competition as a process of rivalry, and emphasise that in the real world, productive and dynamic efficiency and attempts at innovation prevent the market from becoming static, and that this factor is neglected by the standard theory of competition (Vickers, 1995; Blaug, 2001).

The view of competition as a static state initiated with Cournot's (1838/1897) and Bertrand's (1883/1988) seminal oligopoly theories and extended to conjectural variation theory (Bowley, 1924), monopolistic competition theory (Chamberlin, 1933; Robinson, 1933), Stackelberg's (1934/2010) leader-follower model, kinked demand theory (Hall & Hitch, 1939; Sweezy, 1939), and tacit collusion theory (Fellner, 1949; Stigler, 1964). A formal analysis was represented by game theory (Nash, 1950) and repeated game theory (Friedman, 1971; Aumann & Shapley, 1976; Rubinstein, 1979).

15

The roots of the dynamic theories of competition are associated with the acknowledged leaders of the Austrian and Chicago Schools (Mises, 1949; Bork, 1954; Telser, 1960; Shackle, 1971; Kirzner, 1973; Posner, 1979) as well as the insights of contestable markets theory (Baumol, 1982; Baumol et al., 1982) that led to what is now known as New Empirical Industrial Organisation (Pepall et al., 2008).

In the academic literature, the development of several competition theories made the notion of competition appear somewhat complex and hence not directly observable. As a result, economists and scholars have developed numerous methodologies attempting to accurately measure competition and market structure. In general, the academic literature on measuring competition is categorised into two main strands: structural approaches and non-structural approaches.

The first strand, i.e. structural approaches, is based on the traditional Empirical Industrial Organisation (EIO). Early empirical estimates of the degree of competition were based on the structure-conduct-performance (SCP) paradigm

and efficient-structure hypothesis. Formally, these models are measured using various concentration ratios such as: market shares, k -bank concentration ratio (CR_k) and the Herfindahl-Hirschman Index (HHI). However, some authors have doubted the reliability of the EIO models and the associated measures of competition (Shaffer, 2004b; Matthews et al., 2007).

In response to the shortcomings found in structural approaches, a second strand of non-structural approaches emerged. Based on the New Empirical Organisation (NEIO) School, these approaches can be further divided into two major streams. The first generation is based on oligopoly theory and a static model of competition, and includes the Lerner Index (Lerner, 1934), the conjectural variation model (Iwata, 1974), the Panzar-Rosse H-statistic (Rosse & Panzar, 1977; Panzar & Rosse, 1982, 1987), and the Bresnahan-Lau mark-up model (Bresnahan, 1982; Lau, 1982). The second generation of competition measures aimed to capture the dynamics of the market rather than give a static analysis,

16

and includes the persistence of profits (Mueller, 1977, 1986), the Hall-Roeger model (Hall, 1988; Roeger, 1995); the Boone Indicator (Boone, 2008); the competition efficiency frontier (Bolt & Humphrey, 2010), and the lower bound to concentration approach (Tabacco, 2013).

A search of the literature revealed few studies which provide a review of banking competition literature. For instance, Shaffer (1983) have reviewed three non structural approaches: Iwata, Bresnahan-Lau and Panzar-Rosse models. Also, Bikker and Haaf (2002b) have presented a detailed review of ten concentration measures along with the three non-structural approaches. However, none of these previous studies fully assessed the banking competition literature and combined theories, measures and their application to the banking industry. Besides, research to date has not yet provided a comprehensive and up to date review of literature on banking competition. The present study fills this research gap by reviewing a range of academic publications which deal with banking competition and provides a comprehensive framework that combines competition theories and methodologies of measuring banking competition as well as their empirical applications to the banking sector.

This study has three main aims. The first is to give a brief overview of competition theories developed over the years. The second aim is to provide a comprehensive and critical review of the two structural approaches and nine non-structural approaches for measuring competition, highlighting their key strengths and shortcomings. The third aim is to summarise the applications of these measures in the banking sectors of countries worldwide.

Section 2 describes the systematic review process followed to produce this study. Section 3 gives a brief overview of the development of competition theories since the 18th century. Sections 4 and 5 explain in detail the structural and non structural approaches used to measure competition, with illustrations of the strengths and drawbacks of each one, and the practical applications in the banking sector. Finally, Section 6 concludes the study and suggests future trends.

Figure 2.1 below shows a timeline of competition theories and methodologies applied to measure banking competition.

Theoris of Competition**Measures****of Competition**

Originated by Adam Smith (1776) in *The Wealth of Nations*, inspired others for 2 main views of competition
 Structural Approaches (EIO)
 Non-structural Approaches (NEIO)

Competition as a static state Cournot (1838) Oligopoly Theory:
 Competition as a dynamic conception

Non-formal

Formal CRk

1st generation
 (based on oligopoly theory and a static
 2nd generation
 (Dynamic approach of competition)

Quantity-Competition Bertrand

(1883) Oligopoly Theory: Price-Competition
 The Austrian School led by Schumpeter (1934, 1954), Hayek (1941, 1948), Mises (1949), Shackle (1971) & Kirzner (1973)
 Market Power (MP)
 Hypotheses SCP

Efficient Structure (ES) Hypotheses

X

HHI

Market shares
 model of competition)

The Lerner Index Lerner (1934)

Persistence of Profits. Muller (1977, 1986)

Bertrand-Edgeworth Model
 Edgeworth (1897, 1925)

Conjectural Variation
 Bowley (1924)

Monopolistic Competition
 Chamberlin (1933)

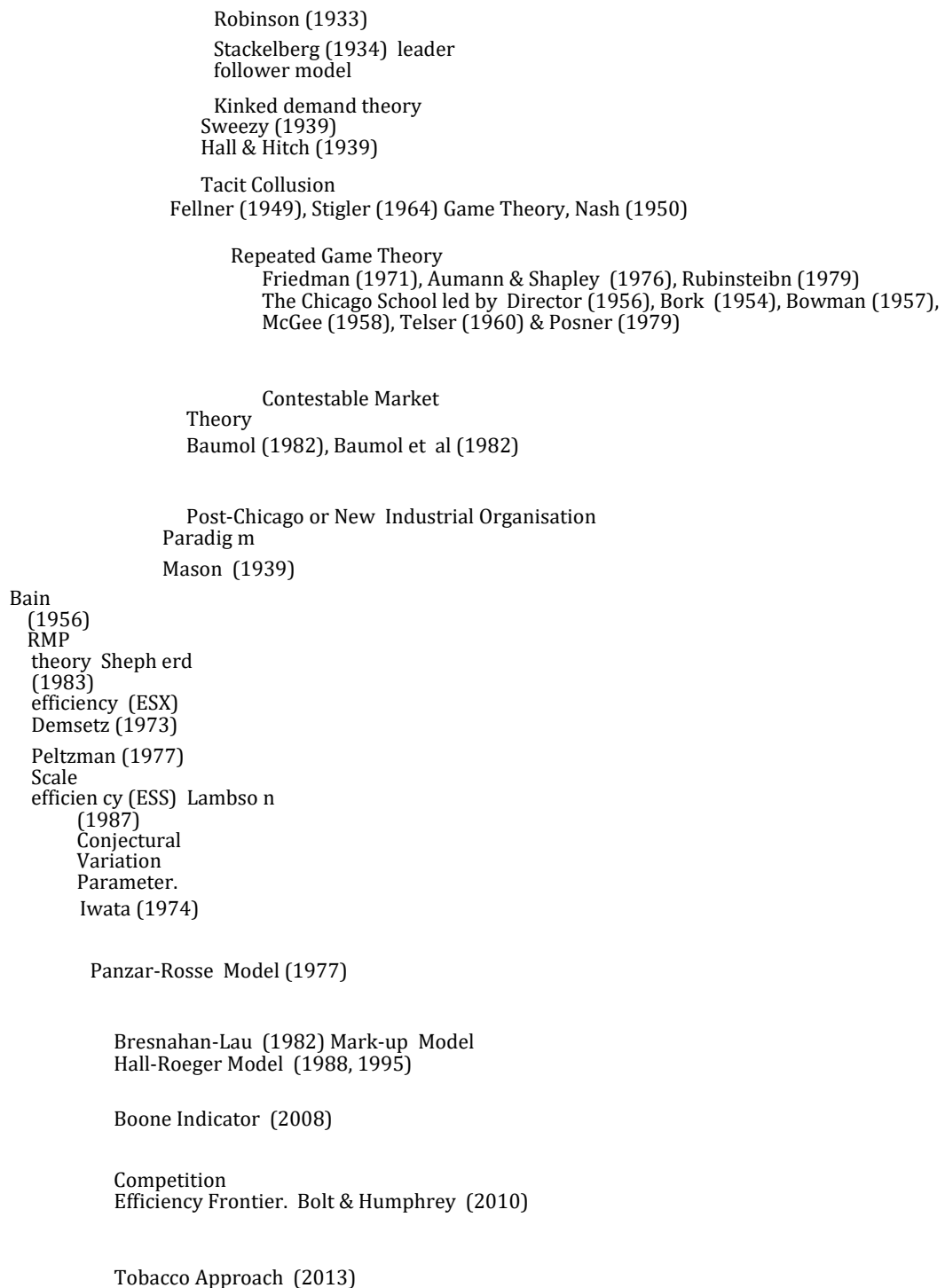


Figure 2.1. Methodologies of measuring banking competition *Source: Author's elaboration*

2. Review Method

The number of published studies on banking competition have an upward trend over the years given the significance and implication of the topic. Figure

2.2 provides a summary of the number of published papers since 1980 in Scopus database.

This study followed the systematic review process proposed by Denyer and Tranfield (2009). The review has focused mainly on published articles, books and book sections and excluded working papers. Also, the review of each competition measure included the first application of the measure in general and then only included its applications in the banking sector. In this vein, this study has analysed a total of 268 references of which 84% is journal articles that were published between 1934 and 2017 in 85 journals (Table 2.1).

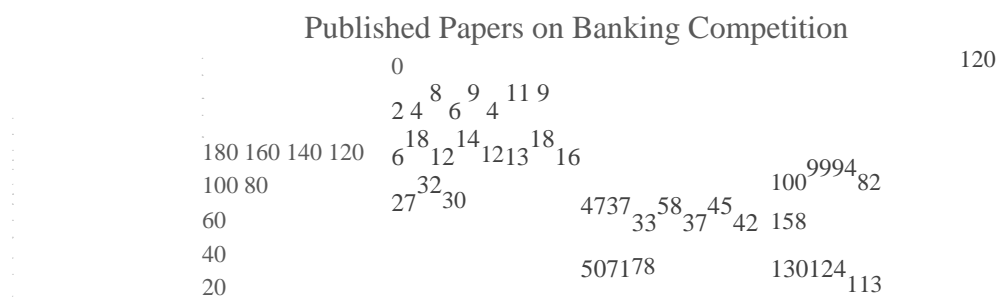


Figure 2.2. The number of published papers on competition in banking (1980-2017) **Source:** Scopus Database

20

Table 2.1. Distribution of articles by knowledge area and journal

Journal

Count Journal Title Article Count
(1934 -2017)

Economics, Econometrics and Statistics

1 American Economic Review 10 2 Applied Economics 9 3 Economics Letters 9 4 Journal of Political Economy 7 5 Review of Industrial Organization 7 6 Journal of Law and Economics 4 7 RAND Journal of Economics 4 8 Review of Economics and Statistics 4 9 Econometrica 3 10 Empirical Economics 3 11 International Journal of Industrial Organization 3
12 Journal of Economics and Business 3 13 Journal of Industrial Economics 3 14 Quarterly Review of Economics and Finance 3 15 Applied Economics Letters 2 16 De Economist 2 17 Economic Journal 2 18 Journal of Agricultural & Food Industrial Organization 2 19 Oxford Economic Papers 2 20 Quarterly Journal of Economics 2 21 Review of Economic Studies 2 22 Review of Financial Economics 2 23 Canadian Journal of Economics 1 24 Economic and Social Review 1 25 Economic Modelling 1 26 Economic Notes 1 27 Economic Systems 1 28 Economica 1 29 European Economic Review 1 30 International Advances in Economic Research 1 31 International Journal of the Economics of Business 1 32 International Review of Applied Economics 1 33 International Review of Economics and Finance 1 34 Journal of Applied Economics 1 35 Journal of Asian Economics 1 36 Journal of Development Economics 1 37 Journal of Economic Literature 1 38 Journal of Economic Methodology 1 39 Journal of Economic Perspectives 1 40 Journal of Economic Theory 1 41 Journal of Productivity Analysis 1 42 Journal of Regulatory Economics 1 43 Journal of the American Statistical Association 1 44 Manchester School 1 45 North American Journal of Economics and Finance 1 46 Scandinavian Journal of Economics 1 47 Studies in Economics and Finance 1 48 Antitrust Bulletin 1 49 Economic and Financial Modelling 1 50 Federal Reserve Bulletin 1 51 International Economics 1 52 Journal of Applied Business and Economics 1 53 Journal of Competition Law and Economics 1 **Finance**
1 Journal of Banking and Finance 36 2 Journal of Money, Credit and Banking 13 3 Journal of International Financial Markets, Institutions and Money 9

21

4 Applied Financial Economics 4 5 Journal of Finance 4 6 Journal of Financial Services Research 3 7 Journal of Financial Stability 3 8 Journal of International Money and Finance 3 9 International Review of Financial Analysis 2 10 Journal of Financial Intermediation 2
11 Research in International Business and Finance 2 12 Review of Finance 2 13 European Journal of Finance 1 14 Finance Research Letters 1 15 Financial Management 1 16 Financial Markets, Institutions and Instruments 1 17 Financial Review 1 18 International Journal of Central Banking 1 19 Journal of Empirical Finance 1 20 Pacific-Basin Finance Journal 1 21 Review of Quantitative Finance and Accounting 1 **Law & Jurisprudence**
1 Northwestern University Law Review 1 2 The University of Chicago Law Review 1 3 The Yale Law Journal 1 4 University of Pennsylvania Law Review 1 **International Business & Area Studies**
1 Emerging Markets Review 3 2 International Business Review 1 **Operational Research & Management Science**
1 European Journal of Operational Research 1 2 Omega 1 **Social Sciences**

3. Competition in Economic Theory

The concept of competition was first articulated by Adam Smith (1776). It was popularised in his book *The Wealth of Nations*, and became the basis of the classical treatment of competition. The well-known Smithian analysis of ‘invisible hand’ principles considers free trade and competition as significant mechanisms to market equilibrium in the long-run. The condition of free trade and competition without government intervention, as if by an invisible hand, would automatically motivate market systems to focus on manufacturing products to serve the nation’s most important needs at prices equal to the costs of production. Smith (1776) also argued that subsidies, monopolies, tariff walls, controls and other privileges that producers enjoy from government authorities threaten free competition and trade and are detrimental. For these reasons,

22

Smith believed that the role of government in society should focus on core functions such as maintaining defence, order, infrastructure and education, while keeping the market and economy open and free to avoid any distortion.

Inspired by Smith, several further models and theories of competition have been proposed which contribute towards two major views of competition. On the one hand, standard theory refers to competition as tranquil or static equilibrium state in which well-informed firms treat prices parametrically and are not able to over-charge and thus make unusually high profits. Under these theories, rivals in the marketplace will compete until equilibrium is reached. Once equilibrium holds in the market, there will be no further interactions between rivals and the market will come to rest. On the other hand, powerful criticisms shed new light on viewing competition as a process of rivalry highlighting the importance of dynamic efficiency and innovation that are neglected in the standard theory of competition. In the real world, efficiency and innovation attempts by market participants do not rest long enough for a static state to hold (Vickers, 1995; Blaug, 2001).

3.1. Static Theories of Competition

Market competitive extremes refers to either monopoly or perfect competition. In both cases, firms do not have to worry about other firms' actions and decisions. In a monopolistic market, there are no other firms, while in perfectly competitive market there are other firms but the decisions of one firm have no effects on other firms. This is because every firm possesses a very small market share and therefore the decisions of that firm would not have any effects on other firms in perfectly competitive market. In the real world, however, the majority of firms and corporations are neither monopolists nor perfect competitors and live in the middle ground of oligopoly where rivals interact strategically to survive (Pepall et al., 2008).

The first precise analytical treatment of static oligopoly models in the literature was pioneered by the French mathematician Augustin Cournot in the 19th

23

century. Cournot (1838/1897) initially examined a monopolistic type of markets and then proceeded to the opposite extreme of competitive equilibrium, which led him to recognise a stage in between, in which a market consists of more than one firm, but not a large number of them. The basic story behind the Cournot model is that a single firm is profitably able to enter a market currently supplied by a monopolist, if certain assumptions are maintained. The entrant firm is assumed to produce identical products and at the same unit cost of the incumbent monopolist as well as free entry and exit conditions in the market. The entrant firm will choose the level of profit-maximising output while taking into account the output level being sold by the monopolist. The former monopolist will hence react and adjust the level of output given the output sold by the new rival firm. The process of choosing the level of output according to that of other firms can be repeated, and represents what Cournot (1838/1897) referred to as reaction curves.

In the Cournot model, the equilibrium market price that will arise upon the firm's choice of production level is less than that of a pure monopoly, yet the equilibrium price is greater than what would prevail in a purely competitive market. Thus, the Cournot (1838/1897) oligopoly theory considers the quantity

of output as the firm's choice or strategic variable when they compete. Although the Cournot model went largely unrecognised for about a hundred years after its publication, it is now considered as a cornerstone of modern industrial organisation theory.

Fifty years later, another influential French mathematician, Joseph Bertrand, introduced the first famous critique to the Cournot oligopoly models. In reviewing Cournot, Bertrand (1883/1988) argued that firms will definitely collude to gain larger profits than the Cournot equilibrium. Moreover, even if firms choose not to collude, they would choose the prices rather than the quantity of produced output. In a situation where firms produce a homogenous output in an oligopolistic market, consumers prefer to buy from the cheapest source. For this reason, it is more reasonable to consider firms as price choosers rather than quantity choosers (Shapiro, 1989).

24

The legacy of Bertrand (1883/1988) is recognising the difference between the strategic variables on which firms compete. According to the Bertrand model, firms will choose and compete at the level of price rather than quantity, and consumers will always choose to buy from the lowest price firm. The Bertrand oligopoly theory also assumes that firms are willing and able to supply all market demand. Therefore, any deviation in price level between rivals will cause an immediate and comprehensive loss of demand for the firm offering higher prices. This gives an incentive to firms to lower their prices in order to capture the whole market. The only Bertrand equilibrium price is then the price sets equal to marginal cost for equally efficient firms with constant marginal cost and homogenous products. This is in contrast to Cournot competition in which prices remain substantially above marginal cost as long as the number of firms is not large. (Shapiro, 1989; Pepall et al., 2008).

Bertrand's criticism of the Cournot model was further elaborated by the Irish philosopher and economist Francis Edgeworth in 1897. Since the original English version of Edgeworth's model of 1897 was lost, the Italian version has been retranslated into English by Edgeworth (1925). One of Edgeworth's criticisms of the Bertrand model was the assumption that firms possess an unlimited capacity for production. While Bertrand's analysis of oligopoly markets assumes that the

only equilibrium for profit-maximising firms with constant marginal cost is the price that equals marginal cost and thus firms earn zero profits, Edgeworth (1925) disagreed for two sound reasons. The first is that rival firms are subject to capacity constraints and are thus unable to serve all consumers in the market. The second reason is that consumers may not view the products of the firms as perfect substitutes. When introducing capacity constraints to the Bertrand model, Edgeworth showed that a pure strategy equilibrium will not generally exist (Dudey, 1992). In the Bertrand-Edgeworth model there is no pure strategy equilibrium and firms play a mixed strategy where prices vary cyclically between low and high (Waldman & Jensen, 2016).

25

Bowley's (1924) introduction of conjectural variations in oligopoly theory adds to the Cournot and Bertrand models. The basic idea is the belief that one firm's choice of output (price) level in an oligopolistic market will affect the rivals' choice of output (price) level. This belief is in contrast to the Cournot and Bertrand model where oligopolistic firms make their choices independently from rivals' reactions to their choices. The reactions of rivals to one firm's choice of either price or quantity is subjectively conjectured by the choice-making firm and hence are called conjectural variations. Those reactions are taken into account and influence the firm's decision when choosing the level of profit-maximising quantity (price). Despite the fact that this theory might refer to a dynamic phenomenon, the analysis is a static one (Bowley, 1924; Dockner, 1992).

An admirable reconciliation to the Cournot and Bertrand oligopoly theories was developed almost simultaneously by the American economist Chamberlin and the British economist Robinson in the year 1933. The starting point of this new departure in theory is that many real-world firms operate under mixed conditions of both market extremes, monopoly and perfect competition. Therefore, an intermediate case in market structure is identified as monopolistic competition. Under monopolistic competition, there are as many firms as in competition, however, with differentiated products, which are not identical but rather very close substitutes. Also, each firm names its own prices in the market. In line with Cournot's and Bertrand's oligopoly theories, the monopolistic

competition theory of Chamberlin and Robinson assumes that decisions and choices of one firm have no noticeable effect on other firms in the market (Chamberlin, 1933; Robinson, 1933; Friedman, 1982).

A step forward in economic theory was proposed in 1934 by the German economist Stackelberg. The duopoly model of Stackelberg (1934/2010) is similar to the Cournot model but is distinguished by a critical difference. According to Stackelberg, both firms are quantity choosers, however the choice of quantity level is conducted sequentially rather than simultaneously. The main feature of the Stackelberg model is the underlying assumption of sequential moves. Stackelberg (1934/2010) thus defined two forms of behaviour, a leader and a

26

follower. The firm that moves first and chooses its level of output is the leader firm, whereas the firm that moves second is the follower firm. In comparison to the Cournot equilibrium, the Stackelberg leader-follower equilibrium yields greater industry output and lower equilibrium prices. Although sequential choice of output can be well-thought-out as a dynamic game, the analysis remains static. The firms are assumed to interact only once and their interaction yields a once and-for-all market-clearing outcome (Stackelberg, 1934/2010; Pepall et al., 2008).

Sweezy (1939), like Stackelberg, presented a special pattern of firms' behaviour and introduced the kinked demand theory in oligopolistic markets. Independently and almost simultaneously, Hall and Hitch (1939) published similar ideas and included more rigorous empirical testing. In the kinked demand theory, oligopolistic firms are presumed to react differently to price changes by one firm in the market depending on the direction of the change. On the one hand, an increase in one firm's prices will be ignored by the rivals and their prices will remain constant. On the other hand, if an oligopolistic firm decides to decrease its prices then other firms would react and decrease their prices by an equal amount. This form of presumed behaviour produces a "kink" in the demand curve of an oligopolistic producer at the existing price. In other words, rivals in oligopoly markets will quickly respond to match price reductions but will not react to price increases. The underlying expectations of firms' behaviour extinguishes the motivation to change prices. The kinked demand theory was

thus an initial attempt to explain sticky prices under oligopolistic market conditions (Hall & Hitch, 1939; Sweezy, 1939; Stigler, 1947).

Oligopoly theories and models developed in the literature stirred the anxiety of economists regarding the stability of equilibria and the tendency of oligopolists to collude. This endorsed the tacit collusion view of Fellner (1949) and Stigler (1964). The tacit collusion theory proposes that firms in an oligopolistic market tend to collude in order to earn maximised joint profits. The level of collusion will depend, however, upon the entry and exit conditions in the market as well as the elasticity of demand.

27

In his influential book, *Competition Among the Few*, Fellner (1949) challenged the usual analysis of supply and demand in an oligopolistic market with a small number of competitors. This challenge stemmed from the fact that the interdependence of actions and the indeterminateness of prices had hitherto largely been ignored. Fellner (1949) also shed light on two main criticisms of the Cournot model. The first criticism is that Cournot was right, but for the wrong reasons, since the stability property makes no sense. In the Cournot analysis, firms would be able to correctly predict their rivals' behaviour and optimise accordingly. Firms will also assume that rivals' choice of output will remain the same once an equilibrium solution has been reached. Fellner (1949), however, strongly rejected this idea and proposed that firms would realise, while they are in the adjustment process toward equilibrium, that their expectations are systematically falsified by evidence and that rivals would constantly change their output level. Therefore, only at equilibrium level are the expectations of firms correct, and once they notice the instability of their rivals' choice, there will be a deviation from this equilibrium. The second criticism is that in an oligopolistic situation with a small number of firms there will be a strong tendency toward collusion. This is a natural phenomenon since the main objective of firms is profit maximisation (Fellner, 1949; Vives, 1989, 2001; Gangopadhyay, 2010).

Stigler (1964) adds to Fellner's argument by stating that secret price cutting by some firms is the major obstacle to the stability of collusive agreements. When a firm observes lower levels of sales, it cannot definitely determine whether it is because of an adverse shock in demand or whether some rivals are cheating on

the cartel agreement. These criticisms emphasise that Cournot's static equilibrium is not valid in a dynamic setting and is a useful notion only when representing the endpoint of a dynamic process. However, the tacit collusion theory lacks the formal analysis of a static Cournot solution concept under dynamic settings (Vives, 1989, 2001; Gangopadhyay, 2010).

In the mid-20th century, an innovative branch of social science emerged in economic theory literature, namely game theory. Game theory represents the

28

first formal analysis and modelling of decisions undertaken in an interactive setting, that is, so-called strategic decisions. A non-surprising aftermath then is that game theory and oligopoly studies are closely intertwined. The unit of analysis divides game theory into two branches: non-cooperative and cooperative. The individual player (firm) is the unit of analysis in non-cooperative games, whereas a group or coalition of players (firms) is the unit of analysis in cooperative games. The underlying application of game theory to oligopoly requires the satisfaction of two basic assumptions. First, firms are rational and pursue well-defined objectives, mainly profit maximisation. Secondly, firms apply their rationality in the process of reasoning strategically. This holds when firms utilise all knowledge to anticipate other firms' behaviour.

In game theory, each firms' actions and decisions are referred to as a strategy. A list of the strategic choices available for each particular player is termed a strategy combination that defines the outcome or final gain of the game. In the context of oligopoly, the game payoff is typically the firms' profits. The equilibrium strategy outcome is the strategy combination where no firm has an incentive to change its strategy given that no other players in the market want to change their strategies. Once such equilibrium strategy outcome holds, it will remain unaltered and the game (market) will come to rest. This type of game in a single period is a static one as firms meet once only and the market then clears once-and-for-all. There are neither sequential moves over time nor repetition of interactions (Nash, 1950, 1951, 1953; Rasmusen, 2007; Pepall et al., 2008).

Nobel laureate John Nash developed this notion of an equilibrium strategy

combination for a non-cooperative game. In his honour, it is commonly referred to as the Nash equilibrium concept. The well-known oligopoly models of Cournot and Bertrand, although developed in the late nineteenth century, incorporate modern game theoretic elements. Cournot's oligopoly equilibrium for a single period quantity model is an extraordinary early example of a non-cooperative game and can be referred to as a Cournot-Nash equilibrium. The same settings apply to the Bertrand oligopoly price model and Stackelberg leader model. Consequently, the equilibrium solutions of those models are the Bertrand-Nash

29

equilibrium and Stackelberg-Nash equilibrium (Friedman, 1982; Pepall et al., 2008).

Game theory has inspired other economists to develop further economic models. This is exemplified in the pioneering work of Friedman (1971), Aumann and Shapley (1976), and Rubinstein (1979) in the theory of repeated games. The main motivation behind this theory is that departing from a static-single period game to a repeated one provides strong incentive for firms to participate in a cooperative-cartel behaviour that would sound dramatically more profitable. Although departing from the cartel agreement is profitable in the short run (one period), the gain will be offset by losses every period thereafter. The theory of repeated game, hence, supports the tacit collusion view and formally analyses the mechanisms to sustain cartels using threats that are credible (Vives, 2001; Pepall et al., 2008). This theory provides an appropriate tool to analyse dynamic interactions rather than single-period interaction; nevertheless, it is referred to as a dynamic analysis of static competition (Audretsch et al., 2001).

The neoclassical concept of competition stems from oligopoly theory. Oligopoly theory classifies the market structure into different categories starting from the collusive extreme, passing by intermediate cases of imperfect competition, and finally reaching the perfect competitive extreme. Oligopoly theory thus inspired economists to derive testable hypotheses and to develop measures of the degree of competition. As a result, the economic literature includes various approaches to gauge the degree of competitiveness. Structural approaches and the early generation of non-structural approaches of measuring competition are based on oligopoly theory and the neoclassical concept of competition (Liu et al., 2013).

Yet, several critiques have been oriented to the static oligopoly theories and the neoclassical conception of competition (Vives, 2001). The main challenge is the contradictory view of competition that focuses on dynamic aspects. The dynamic view of competition refers to competition as being a complex process of rivalry rather than a static concept, and this is discussed in the next section.

3.2. Dynamic Theories of Competition

A controversial view of market structure and competition models questions the usefulness of neoclassical ideas of competition as a static state and their applicability to real world markets. The roots of dynamic models of competition are associated with the acknowledged leaders of the Austrian School of Economics, mainly the early work of Mises, Schumpeter and Hayek as well as a more recent work of Demsetz, Shackle and Kirzner (Audretsch et al., 2001). Principal relevant work includes, among others, Schumpeter (1934, 1954), Hayek (1941, 1948), Mises (1949), Shackle (1971) and Kirzner (1973).

The main argument of the Austrian School economists toward the traditional theories of market competition is the application of competition to a static state of affairs rather than to a process of rivalry. A serious criticism of much of the Austrian School literature on static competition models is that their main focus is about the endpoint of the market, given some initial conditions, after the competitors' interactions have ceased. However, they fail to explain how the market would arrive at this point. In the real world, ongoing entrepreneurship and innovation never stay still long enough for static conditions to hold (Kirzner, 1997; Rosen, 1997; Audretsch et al., 2001). Hébert and Link (1988) add to this argument and further emphasise that static economic theory neglects the role of entrepreneurial innovative actions.

The static economic theories also consider technology and market demand as set, leaving price or output levels as the only variables that firms are able to determine for themselves. The Austrian economists, for example Shackle (1971) and Kirzner (1973), harshly criticise such assumptions and argue that in reality firms constantly seek out and move towards innovation with the aim of gaining

competitive advantage over rivals. Successful entrepreneurs with innovative products would enjoy static market power and earn temporary monopoly profits; yet this kind of competitive advantage will eventually be destroyed by existing rival companies who will imitate the innovation or produce a superior one. Moreover, the few barriers of entry and exit in the industry can also eliminate permanent competitive advantage for successful firms. The prospect of high

31

profits will motivate new competitors to enter the market until economic profits are normalised (Darity & Williams, 1985; Rosen, 1997; Audretsch et al., 2001).

The Austrian prescription for policymakers is that competition authorities should not intervene and impose antitrust regulations if the market enjoys trivial entry barriers. Such intervention to reduce the profits of a successful firm will destroy its incentives to innovate. The only case in which competition authorities should establish regulations and prevent antitrust behaviour is where an incumbent firm in the market has high monopoly power but low levels of innovation. Even in this case, the Austrian prescription suggests that the intervention of competition authorities is still not really necessary, since the lure of high profits will attract new competitors to the market who will defend their entry with the weapon of innovation (Hayek, 1941, 1948; Mises, 1949; Shackle, 1971; Kirzner, 1973, 1997; Audretsch et al., 2001).

The thoughts of the Austrian School economists were revived by Chicago School leaders. The key ideas of the Chicago School antitrust analysis were formulated by Aaron Director in the 1950s and published formally in Director and Levi (1956). The key ideas of Director were further elaborated by lawyers and economists from the school such as Bork (1954), Bowman (1957), McGee (1958), Telser (1960) and Posner (1979). The Chicago view of market competition stresses that contestable markets need to be left without regulations as market mechanisms will naturally abolish any market power and monopoly profits. Chicago School economists argue that firms' actions toward achieving positive profits that are viewed as harmful to competition are actually improving economic efficiency and increasing consumer welfare. The presence of positive profits in a contestable market will then motivate the entry of new competitors until profits cease. Most markets will thus eventually reach the desirable

competitive equilibrium without any regulation intervention.

Both the Austrian School and the Chicago School antitrust analysis shed the light on a *laissez faire* approach to regulation. This view is likewise supported by Baumol (1982) as well as Baumol et al. (1982) in the well-known theory of

32

contestable markets. In line with the Austrian and Chicago Schools, Baumol (1982) and Baumol et al. (1982) defend monopolisation and highlight the importance of entry threat on restraining any monopoly power, which will lead to achieving an eventual maximum social welfare. The contestability theory shows that the incumbent firm is under high competitive pressure from potential rivals who would effortlessly enter the market as long as there are no sunk costs (entry barriers). In perfectly contestable markets, with no barrier to entry and exit, the threat of new entrants will thus discourage any intention of monopoly firms to raise prices.

Subsequently, the conclusions of the Austrian School, Chicago School and the contestability theory have been criticised by several economists. A key critique is questioning the ability of new entrants to compete effectively. Another key critique doubts the common assumption of Austrian economists that entrepreneurial innovative activity, which includes increased costs and risk taking, is always welfare-enhancing. Nevertheless, a significant portion of Austrian thinking is the main root of the new industrial economics that incorporates dynamics and proceeds beyond the confines of static models (Audretsch et al., 2001). The Chicago School contributions also are difficult to underestimate, and their legal influence has spread over many antitrust policies. As a result, the field of industrial organisation has been transformed to reflect what some call 'post-Chicago', or what others simply refer to as New Industrial Organisation (Pepall et al., 2008).

New industrial economics focuses mainly on firms' behaviour in terms of strategy and conduct rather than market structure. According to the new industrial economics, the dilemma for competition policy is that government intervention is necessary depending upon the firm's strategy or behaviour. The influence can be obviously seen in the current antitrust policies and mergers

guidelines in many countries around the world including, among others, the United States (US), the United Kingdom (UK) and the European Union (EU) countries (Pepall et al., 2008).

33

The competition models, theories and views developed in the economic literature inspired scholars and economists to construct measures of competition and market structure. The measures of competition applied in the banking sector are classified into two main categories: structural and non structural.

4. Structural Approaches (EIO Theory)

The structural approaches to competition intensity measurement are based on traditional empirical industrial organisation theory. The term 'structural' stems from the explicit use of market structure to measure the degree of competition. Structural approaches consider market structure as an exogenous variable, and typically regress the profitability of firms on a measure of concentration and a number of control variables. Positive relationship indicates firms' exploitation of market competition and a lack of competition (Matthews et al., 2007). In the economics literature, structural approaches are introduced informally in the form of theories and hypotheses, and formally through the measures of concentration (Bikker & Haaf, 2002b).

4.1. Non-formal Approaches

A theoretical basis for a systematic connotation between firms' profitability and measures of market structure, either concentration or market shares, arises from two market power (MP) hypotheses and two efficient-structure (ES) hypotheses with roots in the traditional empirical industrial organisation (EIO) theory.

The market power hypotheses include two related theories. The Structure Conduct-Performance (SCP) Paradigm of Mason (1939) and Bain (1956) predicts a positive relationship between market structure and firms' performance. According to the SCP theory, this prediction reflects the setting of collusive prices that are less favourable to consumers (lower deposit rates, higher loan rates) as

a result of competitive imperfections in highly concentrated markets. Market concentration would then enable the exploitation of firms' market power and ease collusive agreements to set higher prices and hence enjoy higher

34

profitability (Mason, 1939; Bain, 1956; Shaffer, 2004). A related theory is the Relative-Market-Power (RMP) hypothesis, which argues that only firms with large market shares and well-differentiated products are able to exercise market power in pricing and earn monopoly profits (Shepherd, 1983).

In contrast to the two MP theories, two efficiency explanations of the profit structure positive relationship emerge from the ES hypothesis. The Efficient Structure (ES) hypothesis demonstrates that this positive relationship is defined by the efficiency level of a firm. Demsetz (1973) and Peltzman (1977) claim that firms with creative management and/or superior production technologies usually incur lower costs and are therefore able to generate higher profitability compared to their inefficient counterparts. The ES hypothesis has usually been proposed in the literature in two versions postulating similar ideas. The X efficiency version of the efficient-structure hypothesis (ESX) proposes that firms with higher efficiency have lower costs and will enjoy higher profits, thus generating a larger market share which may result in high concentration levels due to superior management and technology (Demsetz, 1973; Peltzman, 1977). The Scale-Efficiency version of the efficient-structure hypothesis (ESS) suggests that firms have equivalently good management and technologies; however, some firms are more scale-efficient and would produce the output closer to the minimum average cost point. Such scale-efficient firms are assumed to have large market shares that would result in high levels of market concentration. Both hypotheses shed light on efficiency as the main driver of profits and market structure, and point to the falsity of the direct positive correlation between firms' profits and market concentration as illustrated by the two MP hypotheses (Lambson, 1987; Berger, 1995; Al-Muharrami & Matthews, 2009).

There is a large volume of published studies testing the market-power (MP) and efficient-structure (ES) hypotheses in the banking industry worldwide. Some studies have tested the MP hypotheses alone by examining the price concentration association. Reviews of such empirical investigations are

provided by Rhoades (1982), Gilbert (1984) and Weiss (1989). While the literature of concentration-price relationship in the banking industry supports the SCP theory

35

in many cases (Berger & Hannan, 1989; Hannan & Berger, 1991; Bikker & Haaf, 2002; Pilloff & Rhoades, 2002; Al-Muharrami & Matthews, 2009), published results have been mixed.

Pricing conduct has often shown no significant relationship with market concentration, even in the earliest studies (Flehsig, 1965; Meyer, 1967). Rhoades (1982) expresses doubts in the overall ability to link concentration and performance. In a similar model, a comprehensive review of 47 datasets by Weiss (1989) showed that a significant positive association between loan interest rates and market concentration existed in only 45% of the sample reviewed. Similarly, Petersen and Rajan (1995) note that posted interest rates on bank loans are less expensive for new and credit-constrained firms in concentrated banking markets than in unconcentrated banking markets. Less robust evidence and criticisms for the positive correlation between market structure and banks' performance have also been embodied by several lines of empirical research in the literature (Jackson, 1992; Rhoades, 1995; Hannan, 1997; Santomero, 1999).

Historically, research investigating the positive profit-structure relationship have been unable to distinguish among the MP and ES hypotheses, as efficiency was not directly measured. Early examples of research into ES hypotheses regress the profitability of a firm on concentration and market shares. Similar results have been found, however with very different interpretations because efficiency variables, mainly X-efficiency or scale-efficiency, were excluded from the analysis. Some authors argue that a result of positive-significant coefficient estimate of market share and an insignificant coefficient of concentration supports the RMP theory that relates market share to market power (Kurtz & Rhoades, 1992; Rhoades, 1985; Shepherd, 1986a, 1986b). Other authors conversely argue that this finding validates the ESX hypothesis, since there is a possible positive relationship between market shares and X-efficiency under the ESX hypothesis even with the absence of any direct X-efficiency measure in the equation (Smirlock et al., 1984, 1986; Smirlock, 1985; Evanoff & Fortier, 1988).

In the same vein, studies supporting the ESS hypothesis, under which scale efficiencies are positively related to profits and market shares, assume that

36

market share may pick up the correlation with excluded scale-efficiency (Gale & Branch, 1982; Smirlock et al., 1984; Stevens, 1990).

A seminal study by Berger (1995) tested the two MP and two ES hypotheses together and was the first to add direct measures of both X-inefficiency and scale efficiency to the empirical analysis. Despite the limited support found for the ESX and RMP theories, Berger (1995) concludes that all four hypotheses do not appear to be of great importance in explaining banks' profits as efficiency and market power tend to explain relatively small variance of firms' profitability. A more recent study by Homma, Tsutsui, and Uchida (2014) has proposed a new test for the ES hypothesis that directly examines firms' efficiency-growth relationship. Consistent with the ES hypothesis, more efficient Japanese banks seem to defeat competition and grow as a result, obtaining larger market shares.

4.2. Formal Approaches

Numerous empirical investigation of structural competition models have used measures of concentration as a proxy to capture structural features of the market and to explain competitive performance in the banking industry (A. Berger et al., 2004; Beck et al., 2006; Alegria & Schaeck, 2008). According to Hall and Tideman (1967), concentration measures should be a one-dimensional measure ranging between zero and one and be independent of the size of the industry. Marfels (1971), in addition, differentiates between four classification of concentration measures according to their weighting schemes and structure. The first classification is the weights of unity and zero attached to banks in the market based on their market shares. A weight of one is attached to an arbitrarily determined number of banks ranked in descending order, whereas a weight of zero is attached to the remaining banks in the industry. The second classification uses the market share of a bank as its weight, thus allowing greater weights to larger banks. The ranking of individual banks, either descending or ascending, are used as their own weight in the third classification. Finally, some concentration measures weight each bank by using the negative logarithm of its

market shares. This classification therefore gives a smaller absolute value to banks with larger market shares.

Surveying the literature in this field, several concentration measures have been used to capture the degree of competitiveness in many banking markets, such as: number of banks, k -bank concentration ratio (CR_k), Herfindahl-Hirschman Index (HHI), Hall-Tideman Index (HTI), Rosenbluth Index (RI), Comprehensive Industrial Concentration Index (CCI), Hannah and Kay Index (HKI), the U Index (U), multiplicative Huse Index (H_m), additive Huse Index (H_a), and Entropy measure (E) (Bikker & Haaf, 2002b)¹. According to 73 US Structure-Conduct Performance (SCP) studies in banking, the most frequently applied concentration measure is the k -bank concentration ratio, and particularly the 3-bank concentration ratio, followed by HHI and the number of banks (Phil Molyneux et al., 1994). Simplicity and limited data requirements are salient features of both k -bank concentration ratio and the HHI. Therefore, these measures only will be explained in detail.

The k -bank concentration (CR_k) ratio can be simply measured by summing the market shares of k banks in the market. Algebraically, it is measured as:

$$CR_k = \frac{\sum_{i=1}^k MS_i}{n}$$

where MS is the market share of the i th bank and k is the number of the largest banks included in the calculations of the ratio. The index values range between zero and unity. Zero value of the index indicates an infinite number of equally sized banks, while unity is reached if the entire market is made up of the banks included in the calculations of the ratio. A distinguishable aspect of the CR_k index is that it does not discriminate between the k leading banks. However, it neglects other smaller banks in the market and the decision of determining the value of k is somewhat arbitrary (Al-Muharrami et al., 2006; Bikker & Spierdijk, 2010).

¹ Bikker and Haaf (2002b) provide detailed information on the calculation and interpretation of the mentioned measures of concentration.

Index (HHI) independently. The HHI Index has been the basis of merger guidelines in the US Department of Justice since 1982 and is widely used by bank regulatory agencies. The HHI is a static measure of concentration at a single point of time and is calculated by squaring the market share of each bank in a particular banking market and then summing the squares, as follows:

$$HHI = \sum_{i=1}^n MS_i^2$$

As mentioned above, MS refers to the market share of the i th firm, while n refers to the number of firms in the market. The HHI values can range from zero in a market with an endless number of banks, to 10,000 in a market with only one firm. According to the 2010 Horizontal Merger Guidelines, the markets are classified into three types: unconcentrated markets ($HHI < 1500$); moderately concentrated markets ($1500 < HHI < 2500$); and highly concentrated markets ($HHI > 2500$). In general, mergers in a moderately and highly concentrated market involving an increase in the $HHI > 100$ points are likely to raise substantial competitive concerns and often warrant scrutiny (U.S. Department of Justice and The Federal Trade Commission, 2010). Unlike the k -bank concentration ratio, the HHI incorporates all banks in the market and thus arbitrary decisions are avoided. However, it discriminates between banks as it assigns greater weight to larger banks in the market (Matthews et al., 2007; Al Muharrami, 2008; Daley & Matthews, 2012).

In the same vein, Sutton (1991) proposes the bound approach to model the market size–market structure relationship and distinguish the equilibrium structure between the two types of industries. The relationship between the market size and market concentration is traditionally thought to be inversely related (see e.g. Dick, 2007). However, Sutton (1991) argues that this inverse size–structure relationship is valid only for certain groups of industries, while it breaks down for industries in which shifting consumer demand relies on the fixed rather than variable costs. The empirical estimation of the relationship plots a lower bound to concentration. For exogenous sunk costs industry, the

lower bound would be sharply decreasing and eventually approaches zero with limiting level of concentration below 1%. Such industries are of competitive

nature where escalating profits due to market expansion induces entry to the market and hence lowers the concentration level. However, the estimated bound is remarkably flat for endogenous sunk costs industries. This would suggest that number of entrants is entirely unaffected by the size of market and the industry is a natural oligopoly.

As far as theory goes, the entry to such industries is naturally limited to a handful of firms, when quality improvement and shifting the technological frontier outward is predominantly due to endogenously determined fixed costs. A continuous increase in the level of fixed costs is not always feasible for all firms. Therefore, it is believed that there exists a lower bound to the equilibrium level of concentration as the number of entrants is unrelated to the market size, and hence the market fails to fragment. Conversely, as the market size increases, the current incumbents will escalate their fixed investments. As a result, larger markets will not have a greater number of firms; instead the theory predicts that there will be an improvement in the quality of products.

Empirically, the Sutton's (1991) model uses a cross-sectional data and applies either maximum likelihood or minimum distance estimator involving the error term following the extreme-value Weibull distribution as:

$$L_{it} = -(\alpha + \beta \ln((M_t/2^\alpha) + 3^\alpha))$$

where L_{it} is a logit transformation of the relevant concentration ratio for country i at time t and $3^\alpha > 0$. (M_t corresponds to the market size and 2^α is a proxy of set-up cost that captures the minimal sunk cost required by each entrant to the industry prior to establishing production.

Although Sutton's theory is developed in the early nineties and provides solid predictions of the size-structure relationship across a broad class of competition models, surprisingly, the empirical applications of Sutton's theory has been remarkably sparse. Ellickson (2007) is one of few exceptions. He was the first to

test the theory on a large sample of markets within a single industry. His work examines the theory's predictions on the US supermarket industry, where a few

firms were found to dominate the market and capture the majority of sales regardless of the market size. Also, the estimated lower bound was remarkably flat, indicating that the market concentration is unaffected by the market size.

The empirical testing of Sutton's theory requires defining a proxy of the set-up costs, which represent the minimal level of sunk cost required for anyone to enter the industry. However, this information is not available for most industries or is difficult to estimate. For instance, Dick (2007) applies Sutton framework to the US banking industry but the market size was not scaled by the set-up ratio as it appears to be that there is no legal minimum. Inasmuch as the Sutton analysis is powerful to make prediction about the relationship between market size and market concentration, the empirical analysis requires data on entry requirements, which are not easy to come by. The limited empirical research testing of Sutton's theory owes to failure to obtain these data because it either does not exist at the micro level or cannot be disclosed due to confidentiality.

Several critics question the ability of structural approaches to capture the competitive conduct of a market. One criticism of much of the literature on the use of structural approaches as a proxy of competition centres around the role of market contestability. Contestability theory, as initially advanced by Baumol (1982) and Baumol et al. (1982), claims that alternate market conditions of no sunk costs and possibility of hit-and-run entry can undermine the degree of competition in a market. Another important criticism is advanced by Claessens and Laeven (2004), who find no empirical evidence supporting the inverse concentration-competition relationship. Even more surprisingly, significant market power is found in highly unconcentrated markets. This assertion is further supported by Schaeck et al. (2009), who examined competition conduct in 45 banking systems worldwide. The investigation results show that concentration and competition describe different characteristics of banking markets.

Likewise, a considerable amount of empirical banking literature has shown that concentration is a poor measure of competition (Berger et al., 2004; Shaffer, 1993, 1999, 2002a; Shaffer & DiSalvo, 1994). Some of these studies have found

banks' behaviour considerably more competitive than proposed by the market structure, whereas other studies have found significantly higher market power than would be implied by the market structure. Since the mismatch can keep running in either directions, structural approaches are tremendously unreliable measure of performance (Bikker et al., 2012).

The existing pitfalls and limitations in the structural approaches have led to the embodiment of non-structural approaches that rely on comparative statistics of profit-maximising firms.

5. Non-structural Approaches (NEIO Theory)

The various limitations of structural approaches are refined by New Empirical Industrial Organisation (NEIO) techniques, which measure the degree of competition directly without explicitly using any market structure indicator. NEIO techniques include a variety of methodologies requiring different data and assumptions to measure the degree of market power in an industry. Compared to the structural measures, the main advantage of these approaches is that they focus on the deviation of output price from marginal cost rather than profitability, and therefore offer a more robust theoretical base on which to estimate competitive conditions in the banking sector. These approaches are distinguished by measuring competition without the explicit use of a market structure indicator, hence are called 'non-structural' (Matthews et al., 2007; Beck et al., 2013). Non-structural approaches can be further classified into two generations.

The first generation of non-structural approaches is based on the oligopoly theory or the so-called 'neoclassical conception of competition' as well as a static model of competition. These approaches include the Lerner Index (Lerner, 1934), the conjectural variation model (Iwata, 1974), the Panzar-Rosse H statistic (Rosse & Panzar, 1977; Panzar & Rosse, 1982, 1987), and the Bresnahan-

Lau mark-up model (Bresnahan, 1982; Lau, 1982). Although the first generation of non-structural approaches generally share a common standard theoretical framework, the models often draw divergent conclusions. The second generation of non-structural approaches, conversely, focuses on capturing the dynamics of a

market rather than a static condition and is therefore in line with the Austrian dynamic concept of competition. These include the persistence of profits model (Muller, 1977, 1986), the Hall-Roeger model (Hall, 1988; Roeger, 1995); the Boone Indicator (Boone, 2008); the competition efficiency frontier (Bolt & Humphrey, 2010), and the lower bound to concentration approach recently proposed by Tabacco (2013).

In most of the NEIO techniques, banks' inputs and outputs have to be determined via a banking firm model. Three fundamental models have been identified in the literature. Under the intermediation model developed by Klein (1971) and Sealey and Lindley (1977), the production function of banks employs labour and physical capital to generate deposits, which in turn are transferred into funding loans and other earning assets. The model thus identifies three main inputs that banks use to generate their outputs: price of deposits (ratio of interest expense to total deposits), price of labour (ratio of wage expenses to the number of personnel), and price of physical capital (ratio of total expenses on fixed assets to the total value of fixed assets).

The other two main models apply comparative definitions with two vital refinements. In the value-added model (A. Berger & Humphrey, 1992), deposits are classified as outputs rather than inputs. The core idea behind this classification is that deposits provide depositors with valuable services in terms of safe storage of value, record keeping and as a means of payment for some type of deposits. On the other hand, the user-cost model (Hancock, 1985, 1991) applies a different categorisation of bank inputs and outputs. Demand deposits act as bank outputs, given the positive net revenue earned on these accounts on average, whereas other types of deposits act as bank inputs due the fact that banks have to pay a positive net user cost to attract and maintain those funds. Further empirical evidence, however, recommends that deposits overall behave

There is a large debate in the academic literature concerning the reliability of the different competition measures. As a matter of fact, the literature on measuring

market power in banking industries has revealed several contrasting themes, since different research methodologies yield different findings regarding the market structure and competitive conditions even when applied to the same banking market. While some authors prefer one methodology over another, all methodologies have both advantages and drawbacks. The application of a specific methodology is usually constrained by data availability and complexity of econometric techniques needed (Claessens & Laeven, 2004; Carbó et al., 2009; Bikker & Spierdijk, 2010; Bikker et al., 2012). The vast majority of studies on investigating the degree of competition in banking have utilised the Lerner Index, the Panzar-Rosse H-statistic, the Bresnahan-Lau Model and the Boone indicator, making these measures the most widely applied in the banking literature.

This section focuses on the most widely applied measures of competition mentioned above and will briefly discuss the others while illustrating their main advantages and drawbacks.

5.1. First Generation of Non-structural Approaches

Lerner Index (Lerner, 1934)

The deviation of output price from marginal cost is the essence of market power. In theory, a perfectly competitive firm operates at the point where price equals marginal cost. *Ceteris paribus*, a divergence between these values in the long run indicates the existence of monopoly power. This divergence reflects both allocative and distributive effects of monopoly power, which may be positive or negative in their overall effect on public welfare (Lerner, 1934). The traditional Lerner Index (or price-cost margin) is defined as the ratio of disparity between firm's output price and its marginal cost to the output price. Algebraically, the Lerner Index can be computed as:

$$L_i = \frac{P_i - MC_i}{P_i}$$

where P and MC are firm i 's price and marginal cost respectively. Under standard presumptions, the Lerner Index ranges from zero in highly competitive markets up to the theoretical limit of one, where firms' market power become greater. In the banking framework, the Lerner index captures the mark-up that banks

charge to their clients by ascertaining the distinction between loan interest rates and marginal costs and communicating it as an extent of the previous. *Per se*, the Lerner price-marginal cost index can be considered as a direct measure of competition (Beighley & McCall, 1975; Agoraki et al., 2011; Coccoresse, 2014).

A considerable amount of literature has been published on testing the competitive conditions in the banking systems applying the Lerner Index. These studies are considered relatively recent, given the development of this approach since the mid-1930s. The relatively recent application of the Lerner Index is due to the difficulty associated with the assessment of banks' marginal costs. In one of the earliest studies, Beighley and McCall (1975) applied the Lerner Index to measure the market power of the US commercial banking market over the year 1968. The Lerner measurements indicate that American banks' market power increased with the existence of fewer banks in the market, greater inequalities among banks, and larger market shares.

Angelini and Cetorelli (2003) derived the calculation of the marginal costs from Iwata's (1974) approach and computed the Lerner Index to capture the dynamic evolution of banking competition in Italy between 1983 to 1997. The competitive conditions seem to be stable prior to 1992 and increased substantially after this year, suggesting that the wave of mergers and acquisitions mainly improved Italian banks' efficiency. While banking competition remained unchanged in Spain during 1986-1998 (Carbó Valverdie et al., 2003), a growing trend in market power and concentration can be seen in five European countries during the 1990s: France, Germany, Italy, Spain and the UK (Maudos & Fernández de Guevara, 2004; Fernández de Guevara et al., 2005). Similar results are also found in larger cross-country studies that included 15 European Union banking

markets (Fernández de Guevara et al., 2007; Maudos & Fernández de Guevara, 2007; Carbó et al., 2009).

On the one hand, many researchers have investigated the competitive conditions using the Lerner Index for individual countries banking markets including, among others, Koetter and Poghosyan (2009), Hackethal et al.(2012) and Inklaar

et al. (2015) for Germany; Delis et al. (2017) for the US; Kasman and Kasman (2016) for Turkey; Nguyen et al. (2016) for Vietnam, China and India; as well as Tan (2016) for China. On the other hand, there are a number of large cross country studies which compare the Lerner Indices and market power over a large group of banking industries (Berger et al., 2009; Turk-Ariss, 2010; Agoraki et al., 2011; Coccoresse, 2014; Efthyvoulou & Yildirim, 2014; Mirzaei & Moore, 2014; Clerides et al., 2015; Leroy & Lucotte, 2015; Delis et al., 2016; Nguyen et al., 2016; Kouki & Al-Nasser, 2017).

More recently, Koetter et al. (2012) proposed a simple adjustment to the calculation of the Lerner indices to take into account cost and profit inefficiencies that exist in numerous firms, which are implicitly assumed to be zero in unadjusted Lerner indices. Koetter et al. (2012) argue that adjusted Lerner indices tend to be significantly larger than conventional Lerner indices and trend upward over time. Similarly, Brämer et al. (2013) demonstrate that modified Lerner Indices are more qualified indicators of market power than conventional ones. This view is further supported by Huang et al. (2016) as well as Huang et al. (2018) who jointly estimate the Lerner index and cost efficiency in a single step, using stochastic frontier methodology that appears to be a novel idea in the literature. Like Koetter et al. (2012), the models of Huang et al. (2016) and Huang et al. (2017) produce higher mean values of the Lerner Index and smaller standard deviations in comparison to the conventional approach.

The key advantage with this measure of competition is that it can be applied at the bank level as well as over time. The Lerner Index is therefore capable of capturing various patterns of behaviour within the market at any one time, and also over a period of years. Moreover, geographical market definition is not

required when calculating the Lerner Index, which is another potential advantage in contrast to market concentration measures (Beck et al., 2013). Shaffer (2004a) further supports the Lerner Index by arguing that it proves to be related to other NEIO techniques that are formally derived from profit maximising equilibrium conditions. This is exemplified in the work undertaken by Beck et al. (2013), who show that aggregate Lerner indices are significantly and statistically related with other measures of competition and market

structure, particularly the Panzar-Rosse H-statistic, market share, number of banks, HHI and CR₃. Similarly, Delis (2012) notes a high correlation between the Boon Indicators and the Lerner indices at bank-level for 84 banking systems worldwide. However, although the Lerner Index is characterised by its simplicity of calculation despite a limited number of observations and its straightforward interpretation, several authors have pointed out its shortcomings.

Some authors argue that a major theoretical drawback of the Lerner Index is that it is a measure of pricing market power rather than a proxy of competition as there are other possible scenarios in which price-cost margins increase with more fierce competition (Stiglitz, 1987, 1989; Cairns, 1995; Bulow & Klemperer, 2002). Other authors maintain that Lerner Indices may over-estimate the banks' market power when risk taking is taken into account (Fernández de Guevara et al., 2005; Oliver, Fumás, & Saurina, 2006; Berger et al., 2009). Vives (2008) adds to the above arguments by asserting the inability of the Lerner Index to appropriately measure the degree of product substitutability. As a matter of fact, Boone (2008) and Boone et al. (2013) maintain that individual Lerner Indices, if converted to a weighted average Lerner Index, may not necessarily reflect the average degree of market power due to the reallocation effect from inefficient to efficient firms. Efficient firms have higher price-cost margins than their counterparts. Thus, the weighted average Lerner index can increase if the increase in the market share of more efficient firms overcompensates the decrease in the respective individual Lerner indices.

Conjectural Variation Model (Iwata, 1974)

47

Conjectural variation represents the reaction of one firm to change in market shares and pricing by its rivals and is a key concept in oligopoly theories. Iwata's (1974) pioneer model has proposed an econometric approach to estimate numerical values of conjectural variation for firms supplying homogenous products in an oligopolistic market of a small number of firms serving a large number of consumers. The value of the conjectural variation can be calculated by defining the marginal cost and the price elasticity for each firm. This will be accomplished by estimating the firm's cost function and market demand function

respectively, while maintaining three essential assumptions, i.e. that price elasticity of demand, firm's marginal cost, and conjectural variation are constant for each firm in each period. The value of conjectural variation under this approach must be larger than -1 (Iwata, 1974).

The established model can generally be used as an effective method for an empirical analysis of firms with homogenous products in an oligopolistic market. However, applications of the Iwata (1974) model in the banking industry are very rare and posed difficulties given the lack of necessary micro-data on the structure of cost and production (Bikker & Haaf, 2002). Spiller and Favaro's (1984) study is the first application of the Iwata (1974) model to the banking market. The results indicate that strategic behaviour of incumbent Uruguayan banks is highly affected by threats of entry, whereas conjectures are heterogeneous among banks between 1977 and 1980. Further applications are undertaken by Gelfand and Spiller (1987) and Berg and Kim (1994) for Uruguayan and Norwegian banking industries respectively.

A broader perspective has been adopted by Shaffer and DiSalvo (1994) who combine the conjectural variation model with the Panzar-Rosse H-statistic to empirically test the conduct of a banking duopoly market in Fulton County in south central Pennsylvania over the period 1970-1986. Both methodologies suggest that banks' behaviour appears imperfectly competitive, but far from collusive. To the best of current knowledge, these are the only applications that exist for the Iwata (1974) model in the banking industry.

48

Panzar-Rosse H-statistic (Rosse & Panzar, 1977)

The Panzar-Rosse (PR) H-statistic model was initially developed by Rosse and Panzar (1977) and expanded later by Panzar and Rosse (1982, 1987). Its basic idea is that each bank adopts an altered pricing strategy in response to a change in input prices, depending on the structure of the market in which banks operate. The PR test is based on properties of a reduced form log-linear revenue equation, rather than a structural equation, at the firm or bank level, and uses a test statistic H .

The bank revenue function is estimated, in which bank revenues are explained by input prices, bank-specific variables and macroeconomic variables that affect the banking market as a whole. The H -statistic of the Panzar-Rosse model is then calculated as the sum of the elasticities of total revenue with respect to each of the bank's input prices. Rosse and Panzar (1977) and Panzar and Rosse (1982, 1987) demonstrate that this measure is negative or equal to zero for a neoclassical monopolist or collusive oligopolistic situation, between zero and one for a monopolistic competitor, and equal to unity for a competitive price taking bank in long-run competitive equilibrium. Figure 2.3 summarises the Panzar-Rosse H -statistic calculations and interpretation.

In order to obtain valid results, three assumptions should be satisfied. First, banks must be treated as single-product firms. Second, there should be no correlation between greater input costs and greater quality services that would generate larger revenues, as such correlation may lead to a biased H -statistic

49

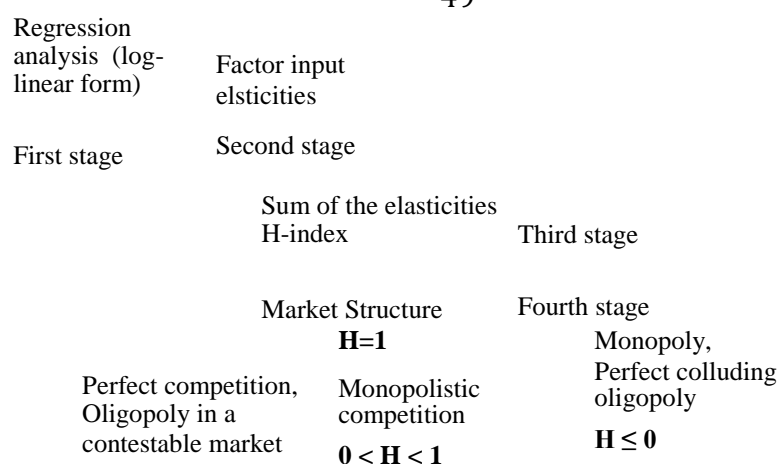


Figure 2.3. Extraction and interpretation of the PR H-statistic.

Source: Apergis, et al., (2016)

value. Lastly, all banks included in the test must operate under long-run equilibrium conditions (De Bandt & Davis, 2000; Bikker & Haaf, 2002a; Yildirim & Philippatos, 2007; Schaeck et al., 2009; Akin et al., 2013).

The pioneering work of Rosse and Panzar (1977) was the first application of a monopoly power test. The H -statistic has been estimated using cross-sectional data for newspaper firms in the local media market (Rosse & Panzar, 1977). In the banking industry, there have been extensive applications of the Panzar and Rosse methodology to assess the competitiveness of the banking sectors in many countries around the world.

One of the first applications of the Panzar-Rosse model in the banking sector is that of Shaffer (1982). For a sample of US banks, the test results showed the existence of monopolistic competition among New York banks in 1979. This application was followed by Nathan and Neave's (1989) study of a sample of Canadian banks. The empirical results reported perfect competition conditions in the Canadian banking market during 1982 and monopolistic competition conditions between the years 1983 and 1984.

50

Using the same model, Molyneux et al. (1994) found a monopoly market structure in Italy, whereas monopolistic competition held for the banking markets of France, Germany, Spain and the United Kingdom. Similarly, monopolistic competition condition has been the market structure across Finnish banks (Vesala, 1995), Japanese banks (Philip Molyneux et al., 1996), Greek banks (Hondroyiannis et al., 1999), South Eastern European banks (Mamatzakis et al., 2005) and major British banks (Matthews et al., 2007).

The case of Italy has yielded contrasting results. While Molyneux et al. (1994)

argue that Italian banks behave monopolistically, Coccoresse (2004) and Trivieri (2007) claim that they operate under monopolistic competition conditions. Even more interestingly, Goddard and Wilson (2009) suggest that the Italian banking market is distinguished by the perfect competition extreme. They also argue that UK banks possess more market power than proposed by Matthews et al. (2007). Updated evidence in favour of monopolistic competition among several European Union banking markets is also found in more recent applications of the Panzar-Rosse model (Delis, 2010; Moch, 2013; Weill, 2013; Apergis et al., 2016).

Research by De Bandt and Davis (2000) in France, Germany, Italy and the US differentiates between types of banks. The empirical findings suggest that small banks enjoy more monopoly power than large ones. This view is supported by Bikker and Haaf (2002a), who find that competition is stronger between large banks and weaker between small banks in the banking markets of industrial countries.

A significant analysis and discussion of the drivers of bank competition using international evidence was presented by Claessens and Laeven (2004). For a sample of 50 countries, the PR model results show monopolistic competition conditions in all of the banking markets examined. Other authors have highlighted the homogeneity in competitive banking markets conditions around the world based on large cross-country comparisons, and many studies have reported monopolistic competitive banking environments (Schaeck et al., 2009; Chen & Liao, 2011; Anginer et al., 2014; Huang et al., 2014).

51

In the mid-2000s, studies investigating degrees of banking market power were extended to emerging economies. Many authors applied the PR model to the banking markets of several Asian and Latin American countries. Similar to the case of most European countries, most of these studies have rejected hypotheses of monopoly and perfect competition supporting monopolistic competition conditions (Perera et al., 2006; Yuan, 2006; Günalp & Çelik, 2006; Yeyati & Micco, 2007; Yildirim & Philippatos, 2007; Park, 2009; Jeon et al., 2011; Olivero et al., 2011; Daley & Matthews, 2012; Liu et al., 2012; Akin et al., 2013; Apergis, 2015; Tabak et al., 2015; Barbosa et al., 2015; Mulyaningsih et al., 2015).

There is a relatively small body of literature that is concerned with examining the market structure and competitive conditions of the Gulf Cooperation Council (GCC) region. One study by Al-Muharrami et al. (2006) examined the trend of competition in the six countries, which are Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and United Arab Emirates (UAE). The PR model outcomes reveal perfectly competitive banking markets in Kuwait, Saudi Arabia and UAE. Conversely, lower degrees of market power and monopolistically competitive behaviour have been noticed in Bahrain and Qatar. The study however states that competitive conditions in Oman should be considered undetermined since the estimated H statistic and hypothesis testing yields contrasting results. A number of other studies have examined the market power in some GCC countries either as a part of Arab Middle East (Murjan & Ruza, 2002), the Middle East Northern Africa (MENA) region (Turk-Ariss, 2009), or as a part of global perspective (Turk-Ariss, 2010).

The PR model has dramatically grown in popularity and is considered to be the most widely applied test for competition in the banking industry with more than 500 journal articles listed in Google Scholar since 2012 alone (Shaffer & Spierdijk, 2015). The popularity of this valuable tool for gauging the degree of competition in banking markets is mainly attributable to its straightforwardness, transparency, and efficiency. Also, information accessibility turns out to be substantially less of a constraint, since bank income is more prone to be

discernible (as opposed to output prices). Additionally, by using bank-level information, this methodology takes into consideration bank-specific contrasts in the production function (Mamatzakis et al., 2005; Trivieri, 2007; Delis, 2010). A critical feature of the Panzar-Rosse methodology is that it measures the monopoly power and analyses the competitive conduct of banks without utilising explicit information about market structure (Bikker & Haaf, 2002a).

However, there are certain drawbacks associated with the use of the Panzar Rosse H -statistic as a measure of market power. One of these is that false indication of market structure can be yielded if the market is not operating in long-run equilibrium. In other words, the risk-adjusted rates of return should be equalised across banks in equilibrium competitive capital

markets and rates of return are uncorrelated with factor prices (Shaffer, 2004b).

Another drawback is shown in a recent study of Bikker et al. (2012). The authors have recognised several situations in which the sign and magnitude of the H statistic do not correspond to the market structure as commonly believed. The study shows that a valid measure of competitive conduct should be assessed by using only an unscaled revenue equation that requires additional information on cost and market equilibrium. Using price equation or scaled revenue function would not yield valid inferences on the degree of market power. This study is complemented by Shaffer and Spierdijk (2015) who show that the H -statistic of the PR model can be positive or negative for any degree of competition. However, Shaffer and Spierdijk (2015) conclude that the Panzar-Rosse revenue test cannot be used as a measure of competitive conditions (neither as a quantitative measure, nor as a one-sided measure). This conclusion has stark implications for the huge body of banking literature, especially those with important policy implications.

Bresnahan-Lau (1982) Mark-up Model

Bresnahan (1982) and Lau (1982) propose a simultaneous equation model that is based on the conjectural variation method of Iwata (1974), but solves the econometric identification problem presented in the Iwata (1974) model

53

through utilising an alternative parametrisation given by the conjectural variation elasticity (Shaffer, 1983; Toolsema, 2002). The key idea behind the empirical model relies on the fact that profit-maximising firms will select prices or quantities such that their perceived marginal revenue equals the marginal cost ($MR = MC$), which coincides with the demand price under perfect competitive equilibrium ($MR = P$). This, however, corresponds to the industry's marginal revenue in extreme collusive conditions ($MR = P$) (Bresnahan, 1982, 1989).

Non-linear simultaneous estimation of a two-equation system, the market demand and supply equations, parametrise the markup of price over estimated marginal cost as an index of market power (μ). The estimated demand equation

can be regarded as first-order local approximation of the true aggregate demand function and the supply equation is based on the profit maximisation assumption. The estimated monopoly power index represents the extent of deviation of the average firm's perceived marginal revenue schedule from the industry demand schedule and therefore captures the degree of market power exercised by firms in the sample (Shaffer, 1989, 1993; Toolsema, 2002).

The conjectural variation elasticity values range between zero and one ($0 \leq \lambda \leq 1$) parallelism to the economic theory that admits only these values under equilibrium conditions. Price taker firms in perfect competitive conditions equate their perceived marginal revenue with the industry demand price. This means that there is no deviation which coincides with the value of $\lambda = 0$. At the opposite extreme, a value of $\lambda = 1$ determines a difference between the firm's perceived marginal revenue and the demand functions. Consequently, firms act as a joint monopoly under perfect collusive conditions. Intermediate values of the market power index ($0 < \lambda < 1$) represent other oligopoly solution concepts of imperfect competition or collusion (Bresnahan, 1982; Suominen, 1994; Shaffer, 1993, 2002; Uchida & Tsutsui, 2005; Coccorese & Pellicchia, 2013).

Besides being a measure of market power, λ serves as a local estimate of the percentage deviation of aggregate output from the competitive equilibrium level of output. In this sense, λ suggests a measure of aggregate excess capacity. If

54

$\lambda < 0$, then the actual output produced seems to be less than the competitive equilibrium level. More interestingly, if $\lambda > 0$, then the actual output level appears to surpass the competitive equilibrium level, though static allocative efficiency requires a marginal cost pricing outcome of $\lambda = 0$ (Shaffer, 1993, 2004b; Gruben & McComb, 2003).

Like any other model, the Bresnahan-Lau mark-up model relies on some established assumptions. The model assumes firms to follow profit-maximising behaviour and risk-neutral attitude. Firms are also assumed to be price takers with regard to their inputs (Bresnahan, 1982; Toolsema, 2002). A fundamental and sufficient condition to econometrically identify λ is that the demand function must not be separable in at least one exogenous variable that is included in the

demand function but excluded from the marginal cost function (Lau, 1982).

In the area of banking, this method was first applied by Shaffer (1989, 1993) to samples of US and Canadian banks respectively. Suominen (1994) extended the Bresnahan-Lau model to the case of two-products model and found fairly intense competitive conduct in the Finnish banking market prior to the deregulation in the mid-1980s. A dynamic version of the model has been developed by Swank (1995) to test the degree of competition in the Dutch loans and savings market. The test results reveal that both markets were more oligopolistic than Cournot competition.

The applications of this technique may yield different conclusions regarding the state of market power even if applied to the same banking market. An example of this is the case of the US banking market, where the results have been mixed across different applications (Shaffer, 1996, 1999; Zardkoohi & Fraser, 1998; Chang et al., 2012). Conversely, a number of applications to the same banking market may suggest similar competitive conditions as in the Italian banking market that tends to operate under imperfect competition (Coccorese, 1998, 2005, 2008, 2009; Coccorese & Pellicchia, 2013). In addition, significant market power and collusive cartel-like conduct have been found in several European banking markets (Neven & Röller, 1999; Shaffer, 2001).

55

Shaffer (2002b) has contributed to the Bresnahan-Lau model in two ways. Firstly, by utilising firm-specific data rather than industry-aggregate data as originally derived and applied. Secondly, by relying on a marginal cost function implied by translog cost function that is more flexible than the original form suggested by Bresnahan (1982). Using this approach, the competitive conduct among Swiss banks seems to vary across ownership structures. The most market power was exhibited by foreign-owned banks, while the least market power was observed in state-owned and mutual banks. Further applications of the Bresnahan-Lau mark-up model includes, among others, Toolsema (2002) for the Dutch consumer credit market, Gruben and McComb (2003) for the Mexican, Canhoto (2004) for the Portuguese, and Uchida and Tsutsui (2005) for the Japanese, as well as Qin and Shaffer (2014) for the Chinese banking market.

This method is particularly useful in studying the level of competition in the banking sector since it assimilates an index of banks' behaviour by gauging their actual deviation from marginal cost pricing and thus it is relevant to public welfare. Another distinguishing feature of the Bresnahan-Lau model is that it does not require any particular definition of banking markets within a country and estimates of the conjectural variation elasticity remain valid and unbiased regardless of the country's banking regulations and the banking market disequilibrium (Shaffer, 2002b, 2004b). This technique also remains a robust and valid measure of conduct even if demand and costs functions are measured imperfectly (Genesove & Mullin, 1998).

An implicit assumption arising from treating the conjectural variation elasticity as a continuous variable is that banks have continuous reaction functions. This parametrisation therefore can represent any form of oligopoly behaviour that arises from either static or dynamic game (Shaffer, 1996). Furthermore, as mentioned earlier, the model overcomes the econometric identification problem in the Iwata (1974) method and is based on industry aggregate data rather than firm-specific data that are usually less available. Finally, the model enjoys the ability to cope with monopsony power without the need to revisit the model and

56

embody a direct relationship with a natural measure of excess capacity (Shaffer, 2001, 2002b, 2004b; Coccoresse, 2009).

The main disadvantage of the Bresnahan-Lau method is that the interpretation of the market power index can reveal anticompetitive biasness if the sample used fails to span complete markets. This is a particular concern for cross-country samples if the researcher estimates the market demand without taking into consideration the correction for cross-border competition (Shaffer, 2001, 2002b). Tabak et al. (2012) point out that the detailed data requirement on demand and costs illustrates a challenge in this model as they are not easily available. Moreover, negative values of the market power index that refer to pricing below marginal cost can result from several causes such as temporary attempts to gain market share and unforeseen loan losses (Shaffer, 1999). Likewise, the estimated market power index embodies the average value of bank behaviour over many years encompassing the sample period. For instance,

a combination of perfect competitive behaviour in some years and imperfect competitive behaviour in other years would generate an average of imperfect competitive behaviour (Shaffer, 2004b; Uchida & Tsutsui, 2005).

5.2. Second Generation of Non-structural Approaches

Persistence of Profits Model (Mueller, 1977, 1986)

A widely held view is that abnormal profits are usually associated with high concentration and entry barriers. The persistence of profits model developed by Muller (1977, 1986) doubts this view by providing an empirical investigation of the dynamics of firm-level profits. The basis of this approach is testing the hypothesis that competition abolishes any high profits rapidly and firms' profit rates will eventually converge toward their long-run equilibrium values. The alternative hypothesis is that a few incumbent firms are either protected by regulations or possess competitive advantage that can prevent imitation and block the entry of new firms. Abnormal profits therefore persist over years and convergence toward long-run equilibrium is either slow or non-existent. The slow convergence rate toward long-run equilibrium indicates stronger profits

57

persistence and greater departure from competitive ideal conditions (Mueller, 1977, 1986).

Thus far, few studies have applied the persistence of profits approach to the banking industry. Among others, Berger et al. (2000) find that persistence of profits differs between US commercial banks located at different levels of performance distribution. The results suggest that competition impediments and informational opacity are the main determinants of persistence. In spite of escalating competition in the European banking markets, there was a significant persistence of abnormal profits during the 1990s (Goddard et al., 2004). Moreover, persistence of profits is positively correlated with ownership concentration in the Italian banking market between 1997-2000 (Agostino et al., 2005).

Knapp et al. (2006) reveal that profits among US banks need about five years to converge towards average industry norms. While Bektas (2007) concludes that

long-run persistency of profits does not exist in the Turkish banking market, the findings of Athanasoglou et al. (2008) conversely propose the existence of long run profitability persistence to a moderate extent among Greek banks. The results further indicate a slight departure of the Greek banking market from perfectly competitive conditions.

A recent study by Goddard et al. (2011) involved large cross-country dynamic panel data set of 19 developed countries and 46 developing countries. The empirical outcomes show a positive relationship between persistence of bank profits and the level of industry concentration as well as the size of entry barriers. On the other hand, persistence of profits is inversely related to growth rate of GDP per capita and the H-statistic, which is a conduct-based measure of competition intensity. Also, in countries with more advanced institutional development and strong external governance mechanisms, competition tends to be stronger and profitability persistence weaker.

58

More recently, Chronopoulos et al. (2015) have concluded that competition indirectly erodes the gain of abnormal profits. Their findings also suggest that the recent financial crises have contributed towards higher levels of profit persistence among US banks. In a different region, Amidu and Harvey (2016) argue that there is a high level of profit persistence and relatively slow convergence in 29 African banking markets.

The dynamic feature of the persistence of profits approach allows the capturing of market dynamics from year to year rather than providing only a snapshot picture of a dynamic competitive process as in static-models-based measures. Monopoly profits gained in one period could dissolve in the following, rendering intervention by government or regulatory authorities unnecessary. However, the application of this approach requires two strong assumptions. First, that entry and exit to the market are sufficiently free to allow the elimination of any abnormal profits quickly by competition. Second, that all firms' profit rates tend to adjust to the same long-run average value. Under a less restrictive version of the model, monopoly profits disappear rapidly and the convergence toward long

run equilibrium may differ between firms. The requirement of such strong assumptions along with complex implementation of the model in highly unstable locations, such as developing countries, are the main reasons for scant applications of the persistence of profits approach to the banking industry (A. Berger et al., 2000; Goddard et al., 2004, 2011; Athanasoglou et al., 2008).

Hall-Roeger Model (Hall, 1988; Roeger, 1995)

The Hall–Roeger approach was originally proposed by Hall (1988) and subsequently extended by Roeger (1995). The basic realisation of Hall (1988) was that in the absence of monopoly power, the traditional Solow residuals should be independent of variation in the log change of output. Hall (1988) also criticised the reliance of Bresnahan-Lau mark-up model on specific functional form assumptions and suggested a non-structural reduced form approach instead. In an extension to the work of Hall (1988), Roeger (1995) contributes to this approach by enhancing the unbiasedness of the estimated market power parameter through the elimination of unobservable productivity shocks. This can

59

be done through observing the difference between production-based (primal) Solow residuals and cost-based (dual) Solow residuals.

According to Hindriks (2005), the Hall–Roeger approach can be classified as one of the NEIO techniques since the estimated equation relates price to marginal cost and thus can be regarded as a supply relation. In the most compact form, the estimated equation of the Hall-Roeger model is: $\Delta \pi^* = @ \Delta A^*$. In the banking context, the left-hand side of the equation $\Delta \pi^*$ is the growth of interest income per unit of capital, while the right-hand side components include $@$ (mark-up of market-power) and ΔA^* (growth rate of input expenses per unit of capital weighted by input cost shares in output). A value of $@ = 1$, indicate that banks operate under perfectly competitive conditions where banks charge a $\pi = \pi^*$. On the other hand, a value of $@ > 1$ suggest that banks exercise market power and charge prices above their marginal costs (Rezitis, 2010; Mirza et al., 2016).

The Hall-Roeger approach utilises a novel idea in the literature of measuring the degree of market power in an industry, since it uses the Solow residuals to

examine the presence of monopoly power. Numerous studies have examined the degree of market power using the Hall-Roeger approach in several sectors of the economy, particularly the manufacturing sector. Examples include Boyle (2004) in Irish manufacturing, Crespi and Gao (2005) in US rice milling, and Wilhelmsson (2006) in the Swedish food industry.

One of the main obstacles, however, in utilising the Hall-Roeger approach is maintaining the assumption of constant return to scale. Deviation from the constant return to scale will bias the estimation of the market power parameter in either direction, upwards or downwards. Until very recently, the application of the Hall-Roeger approach in the banking industry remains a major challenge given the strong assumptions and stringent data requirements.

Up to now, far too little evidence of market power in the banking industry has been found using the Hall-Roeger model. Rezitis (2010) has provided an alternative method for estimating mark-up ratios of the Hall-Roeger approach

60

that does not involve such strong identifying assumptions as inherent in Hall's (1988) and Roeger's (1995) analyses. The study represents the first application of the model in the banking industry. Rezitis (2010) has examined the degree of competitiveness in the Greek banking system during the pre- and post-mergers wave periods in Greece (1995-1998 and 1999-2004) using three different approaches: the Panzar-Rosse H-statistic, the Bresnahan-Lau model, and the Hall-Roeger model. The results of the Bresnahan-Lau and the Hall-Roeger approaches indicate a shift from a competitive to a non-competitive market structure from the pre-merger period to the post-merger period. This study is followed by Mirza et al. (2016) who applied similar methodologies to Pakistani banks. The results show the tendency of banks in Pakistan to behave competitively over the years 2004-2012.

Boone Indicator (Boone, 2008)

Recently, Boone (2008) expanded the current set of non-structural competition measures by proposing an index elaborated from the efficient structure hypothesis that relates firms' performance to differences in efficiency levels. The innovation behind Boone's (2008) work is the measurement of competition

intensity using relative profit differences. The so-called Boone's indicator is based on the assumption that more efficient firms (i.e. firms with lower marginal costs) gain higher profits or market shares, and moreover, that this effect is stronger the greater the competition in the market.

The magnitude of the reallocation effect in the banking system can be empirically examined as: $B^i = C + \beta \ln(\pi^i / c^i)$, where B^i and π^i represents profits and marginal costs of bank i at time t respectively. β is referred as the Boone (2008) indicator. Since marginal costs cannot be observed directly, Boone (2008) suggests the use of average costs as a proxy. In other words, Boone's (2008) model analyses how banks' profits from loans and other earning assets co-vary with average costs of deposits and other borrowed funds, labour and fixed assets (Delis, 2012; Schaeck & Cihák, 2014).

61

Intuitively, the profits of banks with higher efficiency levels in terms of lower marginal costs are expected to increase, yielding negative values of Boone's (2008) indicator ($\beta < 0$). Intensifying the degree of competition in a market raises the profits of more efficient banks relative to their less efficient counterparts. Greater absolute values of β (more negative) denote higher competition and lower market power. The Boone indicator therefore serves as a continuous index of monopoly power (Amidu & Harvey, 2016; Delis, 2012; Leon, 2015).

This methodology can be seen as a robust measure of competition since the reallocation effect is a general feature of intensifying competition. Although different market mechanisms may lead to higher competition (e.g. increase in the number of bank service suppliers due to low entry costs), higher aggressive interaction and relative inefficiencies among banks, the Boone indicator remains a valid measure of competition as long as the reallocation conditions hold (Boone, 2008).

van Leuvensteijn et al. (2011) have investigated the degree of market power in five major European Union countries: France, Germany, Italy, the Netherlands

and Spain as well as the UK, US and Japan, utilising the Boone indicator for the first time in the banking industry. Their research surpasses Boone's (2008) original model in two ways. Firstly, by calculating the marginal costs from a translog cost function rather than using average costs as a proxy. Secondly, by substituting profits with market share as a dependent variable. Overall, the American banks seem to operate under the most competitive loan markets. German and Spanish banks similarly tend to have the most competitive loan markets in Europe, whereas British and French loan markets are the least competitive.

Using this approach, cross-country comparisons have shown heterogeneous degrees of competition and market power among countries under the same time period given the distinct characteristics of the national banking sectors. An example of this is research carried out by Delis (2012) and Clerides et al. (2015),

62

who studied over 84 banking systems worldwide. Similar cases of banking competition heterogeneity are also found within the Latin American and Caribbean banking markets (Tabak et al., 2012; Kasman & Carvalho, 2014), the US and the European banking markets (van Leuvensteijn et al., 2013; Brissimis et al., 2014; Schaeck & Cihák, 2014), as well as emerging economies banking markets (Duygun et al., 2015). Interestingly, African banking regions have witnessed quite similar levels of competition (Amidu & Harvey, 2016; Léon, 2016; Nguyen et al., 2016).

Further applications of Boone's (2008) indicator to the banking sectors include Jeon and Lim (2013) for Korea; Kasman and Kasman (2015) for Turkey; Kar (2016) for ten vibrant microfinance markets: Bangladesh, India, Nepal, Indonesia, the Philippines, Bolivia, Ecuador, Nicaragua, Mexico and Peru; Khan et al. (2016) for five ASEAN countries: Malaysia, Indonesia, Singapore, Philippines and Thailand; as well as Xu et al. (2016) for China.

The Boone indicator has a number of advantages over the traditional structural measures of competition and some non-structural measures of competition. A major benefit of the Boone indicator is the ability to measure competition of different banking market segments and different bank categories, whereas

many well-known market power measures consider the entire banking market only (van Leuvensteijn et al., 2011; Tabak et al., 2012). Another appealing feature of this approach is that it overcomes the shortcomings associated with traditional structural measures of competition, such as the HHI and 3-bank concentration ratio. Unlike concentration measures that captures the outcome of competitive conduct, the Boone indicator is capable of capturing interaction between banks by focusing on conduct. A notable example is the increase in concentration indices in cases of banks failure or mergers that resulted initially from high rivalry conditions. Hence, relying on concentration measures as a proxy of competition may result in misleading inferences (Schaeck & Cihák, 2014; Clerides et al., 2015). The importance of this feature stems from the recent literature conclusions about a very weak relationship between concentration and competition in banking industries (A. Berger et al., 2004; Claessens & Laeven,